SOLID STATE MATERIALS AND DEVICES

Fredrik A. Lindholm, et al

Florida University

Prepared for:

Air Force Cambridge Research Laboratories
Advanced Research Projects Agency

1 December 1973

ARPA Order No. 1060

Program Code No. 2D1

Contractor: University of Florida

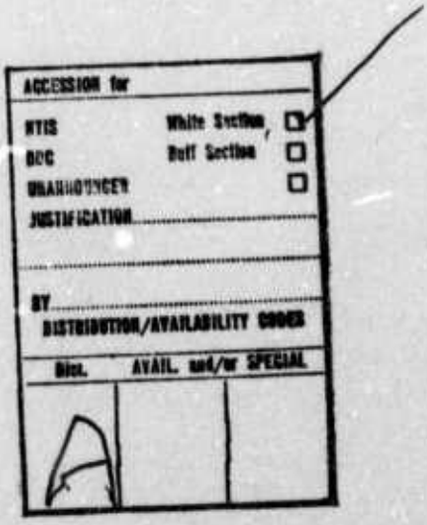
Effective Date on Contract: 1 September 1972

Contract No. F19628-72-C-0368

Principal Investigator and Phone No.
Dr. Fredrik A. Lindholm/904 392-0904

Project Scientist and Phone No.
Dr. Andrew C. Yang/617 861-2225

Contract Expiration Date: 31 August 1



Qualified requestors may obtain additional copies from the Defense Documentation Center. All others should apply to the National Technical Information Service.

it

Unclassified
Security Classification

AD-781165

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

University of Florida
Engineering and Industrial Experiment Station
Gainesville, Florida 32601

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE

SOLID STATE MATERIALS AND DEVICES

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Scientific Final 1 Sept. 1972-31 Aug. 1973 Approved: 20 Feb. 74

5. AUTHOR(S) *(First name, middle initial, last name)*

Fredrik A. Lindholm
Arthur J. Brodersen
Eugene R. Chenette
Sheng S. Li

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
1 December 1973	141 ~~138~~	142

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
F-19628-72-C-0368	
b. PROJECT, TASK, WORK UNIT NOS. 1060 n/a n/a	
c. DOD ELEMENT 61101D	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* AFCRL-TR-74-0044
d. DOD SUBELEMENT n/a	

10. DISTRIBUTION STATEMENT

A - Approved for public release; distribution unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
This research was supported by the Defense Advanced Research Projects Agency	Air Force Cambridge Research Laboratories (LQ) L. G. Hanscom Field Bedford, Massachusetts 01730

13. ABSTRACT

A unification of the circuit models for the large-signal behavior of transistors is given that represents each of the many existing models as a special case of a general model and that unifies the phenomenological inclusion of radiation effects. Questions are examined about the applicability of drift-diffusion transport in the analysis of the behavior of small semiconductor devices; implications concerning transistor speed, saturation current, and temperature dependence receive scrutiny. The effect of neutron radiation on the noise performance of several types of junction and MOS field-effect transistors are studied, and the observations are compared with the predictions of generation-recombination noise theory. The static characteristics, spectral response, and response speed of a new silicon Schottky-barrier phototransistor are measured and compared with theoretical predictions and with the performance of conventional phototransistors. Transient capacitance experiments done for gallium-arsenide Schottky-barrier diodes are used to determine the thermal-emission rate, the doping density, and the thermal-activation energies of oxygen and chromium impurities. Carrier-domain devices are fabricated and evaluated, from both a theoretical and experimental standpoint, with special attention given to a four-quadrant analog multiplier and an arcsine operator.

UNCLASSIFIED
SECURITY CLASSIFICATION

DD FORM 1473-1NOV. 65

DD FORM 1 NOV 65 1473

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
bipolar transistors						
MOS transistors						
junction field-effect transistors						
carrier-domain devices						
radiation effects in semiconductor devices						
photodetectors						
silicon gallium arsenide						

SUMMARY

Two decades of effort have made available many different large-signal transistor models, which provide different compromises between accuracy and mathematical complexity and which contact to different degrees with the processing used in device fabrication. Although certain aspects of various models have been compared, for the most part the relationships among the major existing models remain obscure. The findings reported here seek to simplify, reduce, and unify the results of previous studies to a form that clarifies the relationships among the existing models, the hierarchy of model complexities and accuracies, and the limits of model validity. We demonstrate that all existing models in wide use are special cases of a general model, and give reasons for the belief that future models will likewise be special cases. The general model contains as special cases nearly all existing circuit models for the bipolar transistor: including the original Ebers-Moll model, the charge-control model, the original Linvill lumped model, the modified Ebers-Moll and charge control models developed in connection with CIRCUS, SCEPTRE, NET, SLIC, and other circuit analysis codes, the Gummel-Poon integral charge control model, and the expandable model proposed by Fossum. The general model contains, as well, many circuit models proposed previously for the various types of field-effect transistors: insulated-gate (IGFET or MOSFET), junction-gate (JFET), and metal-gate (MESFET). But the main focus remains on the bipolar transistor. The availability of the general model unifies the phenomenological inclusion of radiation effects; it proves, in principle, to be as easy to include them in any one model as in any other.

Analysis and design of such semiconductor devices as bipolar and field-effect transistors now relies, in part, on the assumption that current flows by drift and diffusion. The minute size of existing devices raises questions about the appropriateness of this assumption, however, because an internal region critical to device performance may become so small that carriers crossing it fail to experience many collisions. To demonstrate possible consequences of continuing to base analysis on

assumed drift and diffusion, we contrast the predictions of two models for the electrical base region of a bipolar transistor. The first model assumes that carriers in transit across the base will experience many collisions, which requires a base region of sufficiently large dimensions. The second model assumes that a carrier in transit will suffer no collisions, displaying what tends to occur in the limiting case of small dimensions. A comparison of saturation current, temperature dependence, and base transit time shows that significant differences exist.

The effect of neutron radiation on noise performance of several types of JFET's and MOSFET's have been studied. Radiation levels which cause very little change in other device characteristics produce drastic increases in low-frequency noise. Radiation-hardened devices showed noise performance consistently inferior to that of good "low-noise" JFET's both before and after irradiation. Measurements have been made of the temperature dependence of the low-frequency spectra.

A new grating-type integrated silicon Schottky-barrier phototransistor has been fabricated and analyzed. The device is constructed by depositing a grating-type aluminum film as a Schottky-barrier contact on the collector-base region of an ordinary phototransistor. The resulting device shows vast overall improvement in responsivity and quantum yield over the corresponding ordinary phototransistor between 0.4 μm and 1.1 μm. With a current gain of 12, the present device has a responsivity of 10.06 μA/μW at 0.6328 μm and 4.1 μA/μW at 0.9 μm, while the corresponding ordinary phototransistor has a responsivity of 2.62 μA/μW at 0.6328 μm and 1.43 μA/μW at 0.9 μm. The bandwidth of the device exceeds 560 MHz and the specific noise equivalent power is found to be 1.93×10^{-14} W at 0.6328 μm. The dc characteristics, the spectral response and the response speed of the present device have been measured and analyzed, and the results are compared with the corresponding ordinary phototransistors.

Transient capacitance experiments have been performed for Au-n type GaAs Schottky barrier diodes between 285 and 316°K. The GaAs substrates are doped with oxygen and chromium respectively. From the transient dark-capacitance and photocapacitance measurements, the thermal emission rate, the doping densities and the thermal activation energies of oxygen and chromium impurities are obtained at T = 300°K.

A new concept in solid-state electronics, the carrier domain, is evaluated. A carrier domain device (CDD) can be fabricated using standard planar diffused silicon processing techniques and exhibits base-emitter and base-collector transfer characteristics which are determined primarily by the planar geometry of the device. A base-emitter structure for carrier domain operation has been designed to serve as a study tool. The two-dimensional base potential and emitter current density distributions in the structure are accurately described in closed-form mathematical expressions over a 95 percent dynamic range of the input signal. The effects of realistic contacts have been accurately accounted for. In the evaluation of the base-collector transfer characteristic, attention is focused upon the motion of the domain centroid which has been theoretically predicted. Two different collector structures have been designed for use with the new base-emitter structure. The two collector designs permit experimental measurement of the motion of the domain centroid along two different coordinate axes. Two CDD's which utilize these base-emitter and collector designs (a four-quadrant analog multiplier and an arcsine operator) have been fabricated and experimentally characterized. The agreement obtained between the theoretically predicted and experimentally measured base-emitter and base-collector transfer characteristics of the devices studied indicates that a macroscopic, two-dimensional representation accurately describes the terminal characteristics of a CDD. It is concluded that either the base-collector or base-emitter transfer characteristic of a CDD can yield acceptable performance in non-linear function generation. It is emphasized, however, that a consideration of the performance of the CDD alone is inadequate. The performance of the complete integrated system into which a CDD is incorporated must be taken into account in the design process.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd.)

TABLE OF CONTENTS (cont'd.)

I. Introduction

The research program sponsored by this contract involves studies in several different problem areas:

(a) the further development of Schottky-barrier photodiodes made with a new mask structure proposed recently at the University of Florida, and collateral studies concerned with material properties relevant to photodetection;

(b) inferences from the noise spectrum measured in semiconductor devices, including irradiated junction field-effect transistors;

(c) modeling of semiconductor devices for computer-determined design including characterization for exposure to radiation environments; and

(d) study of the potential of carrier-domain devices in electronic system applications.

The content of this final report includes treatment of each of these subjects.

II. TRANSISTOR CIRCUIT MODELS (F.A. Lindholm)

Two decades of effort have made available many different large-signal transistor models, which provide different compromises between accuracy and mathematical complexity and which contact to different degrees with the processing used in device fabrication. Although certain aspects of various models have been compared [1-6], for the most part the relationships among the major existing models remain obscure. The circuit designer needs an understanding of these relationships and of the range of model complexity and accuracy available to his use. Such an understanding is needed in the design of an effective and sufficiently general facility for automatic parameter measurement. It is vital in formulating directions for future research.

A main purpose of this report is to simplify, reduce, and unify the results of previous studies to a form that clarifies the relationships among the existing models, the hierarchy of model complexities and accuracies, and the limits of model validity. We shall demonstrate that all existing models are special cases of a general model, and shall give reasons for the belief that future models will likewise be special cases.

The origin of the chaos that persists in transistor modeling, despite intense and well-financed effort, lies in the difficulty of the boundary-value problem describing transistor behavior. That is where the presentation begins. We examine the approximations that have conventionally been made in dealing with this boundary-value problem and thereby are led to the general model and to the physical basis and approximations undergirding it. We then show that the general model contains as special cases nearly all existing circuit models for the bipolar transistor: including the original Ebers-Moll model [7,8], the charge-control model [9,10], the original Linvill lumped model [11], the modified Ebers-Moll and charge control models developed in connection with CIRCUS [12], SCEPTRE [13], NET [14], SLIC [15], and other circuit analysis codes, the Gummel-Poon integral charge control model [16-18], and the expandable model proposed by Fossum [19,20].

We demonstrate that the general model contains, as well, many circuit models proposed previously for the various types of field-effect transistors: insulated-gate (IGFET or MOSFET), junction-gate (JFET), and metal-gate (MESFET). But the main focus remains on the bipolar transistor.

For simplicity, the discussion at first ignores the effects of radiation. We shall see, however, that the availability of the general model unifies the phenomenological inclusion of these effects; it proves, in principle, to be as easy to include them in any one model as in any other.

Certain deficiencies are present in all of the existing circuit models. These contribute to precluding the ability to design an integrated circuit *a priori* by computer. The report ends with a brief identification of these deficiencies, and of their relationship to the exorbitant cost of certain integrated circuits when manufactured in small numbers.

2.1 The Boundary-Value Problem

As Figure 1 suggests, the structure of a bipolar transistor forces variables to depend on three spatial coordinates. Such variables as potential and the densities of current, charge, and mobile carriers--all critical in transistor design--show three-dimensional dependence. Traditionally, device researchers have simplified the resulting three-dimensional boundary-value problem by assuming that practical device geometries permit neglect of dependence in the direction of one of the spatial coordinates. By this assumption the boundary-value problem becomes two-dimensional. As shown in Figure 1(b), one deals with the transistor structure on a per-unit-length basis, similar to that used in treating transmission lines.

Despite this simplification, however, a rigorous computer solution of the relevant equations [21,22] of semiconductor device physics, subject to the boundary conditions imposed by the device geometry and material properties, proves to be difficult. Indeed, no rigorous solutions for time-varying excitation have yet been reported for the two-dimensional boundary-value problem, although progress toward rigor has been made for the dc steady state [23,24].

2.1-1 Sectioning to Further Reduce Dimensions: To avoid the hardships of a two-dimensional solution, one can think of the transistor as comprising two sections: an extrinsic and an intrinsic part, as illustrated in Figure 2. The intrinsic part is mainly the section that makes the transistor useful as an active device; the extrinsic part is largely an unavoidable by-product of the method of fabrication. The intrinsic part lies directly beneath the emitter, which is the control valve in forward-active operation, and extends to a depth below which the mobile electron and hole concentrations resume equilibrium values. The lower boundary shown in Figure 2 is appropriate for transistors that have a buried layer or that are isolated by air or a dielectric. For operation in radiation environments, this class of transistor will find the widest use. The boundary shown is less appropriate for integrated-circuit transistors with junction isolation and without buried layers.

The merit of sectioning as shown in Figure 2 is the resulting simplification of the boundary-value problem. At each point in the intrinsic part, the relevant variables, such as potential and current density, all depend predominantly on a single spatial coordinate.* This coordinate measures the distance from the semiconductor surface. In the extrinsic part, the carrier densities rest approximately at their thermal equilibrium values. For the dc steady state, therefore, modeling of the extrinsic part requires only that one determine values for odd-shaped resistors; if time-varying excitations are applied, the odd-shaped resistors become odd-shaped RC transmission lines, the distributed capacitance being contributed by the transition regions of junctions. Methods exist for estimating values of resistors and capacitors in two-dimensional structures.[25,26].

Thus, sectioning into extrinsic and intrinsic parts reduces a two-dimensional boundary-value problem to a one-dimensional boundary-value problem coupled with side conditions describing the extrinsic part. Computational difficulty is thereby decreased.

Figure 3 displays the two approximations made to this point. In both, the goal is to lessen the number of dimensions in the boundary-value problem. Treating the transistor on a transmission-line basis

---

*Strictly this statement holds only in the absence of current crowding and other phenomena related to multi-dimensional flow.

reduces the number of pertinent spatial coordinates from three to two. This approximation is the better the greater is the aspect ratio of emitter length to width. For high-speed transistors with small and square emitters, it may introduce considerable error. Similarly, though extrinsic-intrinsic sectioning profoundly simplifies the boundary-value problem this approximation too is in question for small devices. Rigorous computer solutions in two dimensions reveal, for certain transistor structures, a large role taken by regions of the base near the emitter sidewall [23]. These effects may possibly be compensated by clever modeling of the extrinsic region.

Despite questions such as these, which remain only partially answered, the treatment in this chapter and in the rest of this report will assume the validity of the two reduction-of-dimension approximations indicated in Figure 3. All of the existing bipolar transistor circuit models derive directly from these approximations. Without exception, derivations of all of these models have tacitly used and accepted them.

In the remainder of this report, we shall focus first upon modeling of the intrinsic part. As we have suggested, the extrinsic part is normally dealt with more easily, and the model for it can be added to give a representation of the overall device.

2.2 General Large-Signal Model for the Intrinsic Part: Static Excitation

Even if the description of the intrinsic part is formulated as a one-dimensional boundary-value problem, a full solution, without using additional approximations, even for dc static conditions, requires use of a computer [27,28]. The solutions resulting from these computer studies give an understanding that aids device design. In tabulated numerical print-out and related graphs, these solutions show the spatial dependence for different bias conditions of potential, charge density, mobile carrier concentrations and other pertinent variables. By linking this dependence with the processing used in fabrication, they lend insight vital to the device designer and provide an alternative to the untenable costs of empirical design. Semiconductor circuits, however, normally contain many devices. For the simulation and design of semiconductor circuits, therefore, direct use of the rigorous computer solutions for the intrinsic part proves to be intractable. The form in

which the information is conveyed is too unwieldy and complex to be used directly as circuit models for transistors. Thus it has been common to make simplifying approximations. By using different sets of approximations, one is led to the various large-signal circuit models now available.

Because the static behavior of a transistor is more easily treated than is the dynamic behavior, it is instructive to focus first on the response to dc static excitation. For the dc steady state, two processes predominate within the intrinsic part: carrier transport between the emitter and collector terminals; and leakage out the base terminal arising from net recombination and injection back into the emitter. The circuit diagram of Figure 4 shows one way of displaying these processes.

All existing circuit models for the large-signal dc behavior of bipolar transistors can be represented by this circuit diagram. This includes the dc versions of all of the models mentioned before: the Ebers-Moll, charge-control, and Linvill lumped models; their generalizations developed in conjunction with the various circuit analysis codes; the Gummel-Poon integral charge control model; and Fossum's expendable Ebers-Moll model. In terms of the circuit diagram of Figure 3, the distinction among these models arises in the functional dependences specified for the current sources. The sources depend on terminal voltages and the constants of device geometry and materials in accord with the specific set of approximations underlying each of the existing static models. The various dependences will be set forth in Section 2.3, which deals with the related models for time-varying response.

2.3 General Large-Signal Model for the Intrinsic Part: Time-varying Excitation

For dc steady-state conditions, the currents flowing into the terminals of the intrinsic part must account for the two processes noted earlier: carrier transport between emitter and collector; and leakage out of the base terminal arising from net recombination and from back injection into the emitter. The circuit diagram of Figure 4 provides one representation of these processes.

For general time-varying conditions, one must consider an additional process: the rate of change with time of the charge q residing within the intrinsic part. Thus

$$i'_E + i'_C = -i'_B = i_{Y1} + i_{Y2} + \frac{dq}{dt} \tag{1}$$

That is, the net current flowing into the intrinsic part from the emitter and collector must balance the current flowing out of the base due to net recombination, back injection, and change in stored charge.* Viewed differently, Eq. (1) states that the net current flowing into the intrinsic part, including displacement current $\frac{dq}{dt}$, must add to zero. Note that $\frac{dq}{dt}$ is the time-rate change of charge of the type constituting i'_E and i'_C. For example, if i'_E and i'_C were wholly constituted of mobile electrons, then $|\frac{dq}{dt}|$ would represent the time-rate change of the charge on the <u>set</u> of mobile electrons within the intrinsic device. In this same example, however, because immense tendencies exist to preserve charge neutrality within the whole intrinsic device, $|\frac{dq}{dt}|$ likewise equals the rate of change of charge carried by the set of mobile holes.

The problem now is to represent Eq. (1) by a circuit diagram. Though there are other alternatives, transport current i_X and current i_{Y1} and i_{Y2} feeding net recombination and back injection can be represented in a way analogous to that used in the dc circuit diagram of Figure 4. The problem thus reduces to finding a circuit representation for the current $\frac{dq}{dt}$.

Although several ways exist to do this [2,26], we concentrate now on the method called quasi-static approximation [2]. Use of this method leads directly to most of the existing large-signal dynamic models and to all of those that have received wide employment in computer simulation of circuit behavior.

This method of approximation supplies not only a circuit representation of $\frac{dq}{dt}$; it furnishes also functional dependences of i_X, i_{Y1}, and i_{Y2} on terminal voltages and on constants of the device geometry and materials.

Quasi-static approximation recognizes that one can find functional dependences for variables in the dc steady state much more easily than for general time-varying conditions. It assumes an intimate relation existing between transient and dc static behavior, and hence the validity of such extrapolations as:

---

*Primes denote variables related to the intrinsic part of the transistor.

$$\text{if } I_X = f_X(V_{BE}, V_{BC}), \text{ then } i_X \simeq f_X(v_{BE}, v_{BC}) \tag{2}$$

$$\text{if } I_{Y1} = f_{Y1}(V_{BE}, V_{BC}), \text{ then } i_{Y1} \simeq f_{Y1}(v_{BE}, v_{BC}) \tag{3}$$

$$\text{if } I_{Y2} = f_{Y2}(V_{BE}, V_{BC}), \text{ then } i_{Y2} \simeq f_{Y2}(v_{BE}, v_{BC}) \tag{4}$$

$$\text{if } Q = f_Q(V_{BE}, V_{BC}), \text{ then } q \simeq f_Q(v_{BE}, v_{BC}) \tag{5}$$

Here standard IEEE notation distinguishes between dc and time-varying quantities: for example, I_X denotes the dc component of current; and i_X the corresponding total instantaneous value.

In these extrapolations, one views transient response as a succession of steady-state responses. Specifically, for the bipolar transistor, one makes the approximation that controlled charge and controlled currents depend on terminal voltages and device make-up in the same way during transients as they do in the dc steady state. For modern devices, in many applications, this approximation probably introduces negligible error. In any case, it underlies most large-signal dynamic models now available for the bipolar transistor. More accurate, though related, methods of approximation have been treated elsewhere [2,29,36].

Combining Eq. (1) with the approximations contained in Eqs. (2) through (5) gives

$$i'_E + i'_C = i'_B = i_{Y1} + i_{Y2} + C'_{BE}\frac{dv'_{BE}}{dt} + C'_{BC}\frac{dv'_{BC}}{dt} \tag{6}$$

in which

$$C'_{BE} = \frac{\partial f_Q}{\partial v_{BE}} \tag{7}$$

and

$$C'_{BC} = \frac{\partial f_Q}{\partial v_{BC}} \tag{8}$$

The circuit diagram of Figure 5 shows one way of representing the physical processes implied in Eq. (6). Topologically, it consists of the circuit diagram of Figure 3, for dc behavior, with two capacitors added.

2.3-1 Matrix Form of the General Model

The functional dependence of each capacitor and current source in Figure 5 depends on the static characterisation chosen in Eqs. (2) through (5). Device theory provides many different static characterizations.

To each different set of static dependences chosen to generate the functional dependences for i_X, i_{Y1}, i_{Y2}, C'_{BE} and C'_{BC}, there corresponds a different large-signal dynamic model. The result is a collection comprising most of the existing models for the bipolar transistor--models having many different levels of accuracy and complexity and many different degrees of contact with fabrication processing.

Figure 5 does not quite represent the most general model; it fails to include a class of models based on higher order-approximations than quasi-static representation of the current $\frac{dq}{dt}$. Without exception, however, all large-signal dynamic models developed to date for the (three-layer) intrinsic transistor are special cases of the matrix equations:

$$\begin{bmatrix} i'_E \\ i'_C \end{bmatrix} = \begin{bmatrix} T'_{11}(s) & -T'_{12}(s) \\ -T'_{21}(s) & T'_{22}(s) \end{bmatrix} \begin{bmatrix} \exp(\theta_{ET} v'_{BE}) - 1 \\ \exp(\theta_{CT} v'_{BC}) - 1 \end{bmatrix} + \sum_{m=1}^{M} \begin{bmatrix} (R'_{11})_m & 0 \\ 0 & (R'_{22})_m \end{bmatrix} \begin{bmatrix} \exp[(\theta_{ER})_m v'_{BE}] - 1 \\ \exp[(\theta_{CR})_m v'_{BC}] - 1 \end{bmatrix} \tag{9}$$

$$i'_E + i'_C + i'_B = 0 \tag{10}$$

Here the primes indicate currents and junction voltages at the terminals of the intrinsic part of the transistor; and s denotes the operator $\frac{d}{dt}$ or, as is sometimes more convenient, the complex frequency variable. The matrix elements T'_{ij} relate to transport of current between intrinsic emitter and collector, and the elements R'_{ij} relate to net recombination-generation mechanisms occurring within the intrinsic transistor. The polarities of the currents and voltages are defined to conform with the standard proposed by Narud and Callahan [30]. For convenience, we will occasionally refer to the square matrices in Eq. (9) as the [T] and [R] matrices.

To derive a particular existing model from Eq. (9), one needs only to specify appropriately $T'_{ij}(s)$, $R'_{ij}(s)$, θ_{ET}, θ_{CT}, θ_{ER}, and θ_{CR}, which describe the elements of the [T] and [R] matrices. To determine the parameters of any existing model, therefore, one needs only to determine, from appropriate data, these same elements of the [T] and [R] matrices.

The most widely used class of models results from so specifying the matrix elements as to produce single-pole approximations of the

frequency dependence of the short-circuit current gains. This specification consists in setting

$$T'_{ij}(s) = T_{ij}(1+s\tau_{ij}), \qquad i = j \tag{11}$$

$$T'_{ij}(s) = T_{ij}, \qquad i \neq j \tag{12}$$

such that

$$[T] = \begin{bmatrix} T_{11}(1+s\tau_{11}) & -T_{12} \\ -T_{21} & T_{22}(1+s\tau_{22}) \end{bmatrix} \tag{13}$$

With these constraints imposed, which cause no important loss in generality, the circuit diagram of Figure 5 represents the general model described by Eqs. (9) and (10).

We now identify the functional dependences of the matrix elements corresponding to each existing model, showing that each is a special case of the general model. We do this first for the single-pole models yielded by quasi-static approximation. Higher-order approximations, which have found much less frequent use in circuit analysis codes, are described in the literature [2,26,31].

2.4 Special Cases of the General Model: The Existing Single-Pole Models

The origin of any transistor model described by matrix equations of the form indicated in Eq. (9) can be traced to Moll's dynamic counterpart [8] of the static Ebers-Moll model [7]. The Beaufoy-Sparkes charge control model [9] and the original Linvill lumped model [11] are mathematically equivalent [1] to the dynamic Ebers-Moll model, as is the nonlinear hybrid-pi model proposed by Howard [32]. The models relating to various network analysis programs--such as CIRCUS, SCEPTRE, NET, CANCER and SLIC -- are straight-forward generalizations. So also is the model proposed recently by Fossum [19,20], though it is more grounded in the underlying physics and hence links more tightly with transistor design. The integral charge-control model due to Gummel, Poon and Chawla [16-18] is another generalization, which also relates to device physics.

2.4-1 General Conditions for Model Equivalence

Figure 6 indicates the topology of the circuit diagram used traditionally to represent the Ebers-Moll model and the models named above which are derived from it. This topology represents both the injection and the transport form of the model [4]. The distinction between the two forms comes from the functional dependences given the current sources and capacitors.

If certain constraints hold, the general circuit diagram of Figure 5 becomes equivalent to the circuit diagram of Figure 6. To demonstrate these constraints, we notice from Figure 5 that

$$i'_E = -i_X - i_{Y1} - C_1 sv'_{BE} \tag{14}$$

$$i'_C = i_X - i_{Y2} - C_2 sv'_{BC} \tag{15}$$

$$i'_B = i_{Y1} + i_{Y2} + C_1 sv'_{BE} + C_2 sv'_{BC} \tag{16}$$

and from Figure 6 that

$$i'_E = -i_{BE} - C_{BE} sv'_{BE} = -(i_{BE} - i_{Y1}) - i_{Y1} - C_{BE} sv'_{BE} \tag{17}$$

$$i'_C = -i_{BC} - C_{BC} sv'_{BC} = -(i_{BC} - i_{Y2}) - i_{Y2} - C_{BC} sv'_{BC} \tag{18}$$

$$i'_B = i_{BE} + i_{BC} + C_{BE} sv'_{BE} + C_{BC} sv'_{BC} \tag{19}$$

Hence the two circuits exhibit equivalent behavior at the terminals if and only if

$$i_X = i_{BE} - i_{Y1} = -(i_{BC} - i_{Y2}) \tag{20}$$

$$C_1 = C_{BE};\ C_2 = C_{BC} \tag{21}$$

Alternatively, these same constraints can be shown to hold in another way. One begins with the circuit of Figure 6. By adding current sources appropriately in pairs so that the net current into each node remains unchanged, one then transforms the topology of Figure 6 into that of the general circuit diagram of Figure 5. The details of this approach are indicated in Figure 7.

Thus, given a model that can be represented by Figure 6, only two conditions need hold to show that this model is a special case of the circuit diagram of Figure 5 and the matrix formulation given by the combination of Eqs. (9) and (13). First, one needs to define the transport current i_X and the capacitors C_1 and C_2 in accord with Eqs. (20) and (21).

This poses no problem. Second, however, for these definitions to have meaning, one needs to be able to express the leakage currents i_{Y1} and i_{Y2} in terms of parameters of the given model. As we shall see, this requirement likewise poses no problem: The physics underlying transistor operation makes plain the appropriate choices for i_{Y1} and i_{Y2}.

We shall now see the unity provided by these results as we consider in turn each of the types of the single-pole models presently available.

2.4-2 Ebers-Moll Model

To derive the dynamic version of the Ebers-Moll model from the general formulation given in Eq. (9), one specifies the slope factors to be the reciprocal of the thermal voltage,

$$\theta_{ET} = \theta_{CT} = \theta_{ER} = \theta_{CR} = e/kT \tag{22}$$

and the [T] and [R] matrices so as to give

$$\begin{bmatrix} i'_E \\ i'_C \end{bmatrix} = I_S \begin{bmatrix} (1 + \frac{1}{\beta_F})(1 + \frac{s}{\omega_F}) & -1 \\ -1 & (1 + \frac{1}{\beta_R})(1 + \frac{s}{\omega_R}) \end{bmatrix} \begin{bmatrix} \exp(\theta_{ET} v'_{BE}) - 1 \\ \exp(\theta_{CT} v'_{BC}) - 1 \end{bmatrix} \tag{23}$$

$$i'_E + i'_B + i'_C = 0$$

which is conventionally represented by the circuit of Figure 8. This is the Ebers-Moll model in its transport form. It is mathematically equivalent to the formulation originally proposed by Moll,

$$\begin{bmatrix} i'_E \\ i'_C \end{bmatrix} = \begin{bmatrix} \frac{I_{EO}(1+s/\omega_F)}{(1-\alpha_F\alpha_R)} & \frac{-\alpha_R I_{CO}}{(1-\alpha_F\alpha_R)} \\ \frac{-\alpha_F I_{EO}}{(1-\alpha_F\alpha_R)} & \frac{I_{CO}(1+s/\omega_R)}{(1-\alpha_F\alpha_R)} \end{bmatrix} \begin{bmatrix} \exp(\theta_{ET} v'_{BE}) - 1 \\ \exp(\theta_{CT} v'_{BC}) - 1 \end{bmatrix} \tag{24}$$

$$i'_E + i'_B + i'_C = 0$$

which is now commonly termed the injection form [4] of the model, also represented in Figure 8. For modern transistors, the transport form, however, gives the more appropriate description [4,16]. In Eqs. (23) and (24), notice that the natural frequencies

$$\omega_F = 1/\tau_{11} \text{ and } \omega_R = 1/\tau_{22} \tag{25}$$

are the reciprocals of time constants, τ_{11} and τ_{22}, defined earlier. The undefined symbols used above have meanings commonly accepted in transistor theory.

Notice that both the injection and the transport forms have the same topology as that indicated in Figure 6. Hence both are special cases of the general circuit diagram of Figure 5, with element values defined by Eqs. (20) and (21). As was discussed in Section 2.4-1, however, to give full meaning to these definitions, one must express the leakage currents i_{Y1} and i_{Y2} in terms of the parameters of the Ebers-Moll model. This one can do by considering the physical process underlying the descriptions of transistor behavior given in Eqs. (23) and (24). Or it can be done more easily by direct examination of the circuit diagrams of Figure 8. By either method, one recognizes that in the transport form of the model the currents

$$i_{Y1} = \frac{1}{\beta_F} i_N \quad \text{and} \quad i_{Y2} = \frac{1}{\beta_R} i_I \tag{26}$$

represent leakage, or recombination. In the injection form

$$i_{Y1} = (1-\alpha_F) i_F \text{ and } i_{Y2} = (1-\alpha_R) i_R \tag{27}$$

are the counterparts.

Thus, from Eqs. (20) and (21), the Ebers-Moll models shown in Figure 8 transform into the general circuit diagram of Figure 5 provided

$$i_X = i_N - i_I = I_s[\exp(\theta_{ET} v'_{BE}) - \exp(\theta_{CT} v'_{BC})] \tag{28}$$

$$i_{Y1} = (I_S/\beta_F)[\exp(\theta_{ER} v'_{BE}) - 1] \tag{29}$$

$$i_{Y2} = (I_S/\beta_R)[\exp(\theta_{CR} v'_{BC}) - 1] \tag{30}$$

$$C_1 = (1 + \frac{1}{\beta_F})\theta_{ET} I_S \tau_{11} \exp(\theta_{ET} v'_{BE}) \tag{31}$$

$$C_2 = (1 + \frac{1}{\beta_R})\theta_{CT} I_S \tau_{22} \exp(\theta_{CT} v'_{BC}) \tag{32}$$

for the transport form of Figure 8(a), and provided

$$i_X = \alpha_F i_F - \alpha_R i_R = \frac{\alpha_F I_{EO}}{1-\alpha_F\alpha_R}\left[\exp(\theta_{ET} v'_{BE}) - \exp(\theta_{CT} v'_{BC})\right] \quad (33)$$

$$i_{Y1} = \frac{(1-\alpha_F) I_{EO}}{1-\alpha_F\alpha_R}\left[\exp(\theta_{ET} v'_{BE})-1\right] \quad (34)$$

$$i_{Y2} = \frac{(1-\alpha_R) I_{CO}}{1-\alpha_F\alpha_R}\left[\exp(\theta_{CT} v'_{BC})-1\right] \quad (35)$$

$$C_1 = \theta_{ET}\tau_{11}\frac{I_{EO}}{1-\alpha_F\alpha_R}\left[\exp(\theta_{ET} v'_{BE})\right] \quad (36)$$

$$C_2 = \theta_{CT}\tau_{22}\frac{I_{CO}}{1-\alpha_F\alpha_R}\left[\exp(\theta_{CT} v'_{BC})\right] \quad (37)$$

for the injection form of Figure 8(b). In the original Ebers-Moll model, all slope factors equal the thermal voltage e/kT. Here, we have employed more general notation to make more concise our later treatment of generalizations of the Ebers-Moll model.

Among the approximations [33] underlying the validity of the Ebers-Moll model several deserve special comment. Ebers and Moll focussed attention on the charge of the minority carriers in the quasi-neutral base, and the function f_Q defined in Eq. (5) is the quasi-static dependence of this charge. This focus of attention results in no explicit account being taken of the Early effect [34] nor of the depletion capacitances at the junctions. These capacitances may be added in parallel with C_1 and C_2, however, which is equivalent to adding terms with appropriate dependences [2,35-37] to Eqs. (31), (32), (36) and (37) above. Additionally, the model assumes low injection of carriers, constant lifetime, negligible net recombination in transition regions, the absence of breakdown and one-dimensional flow. The Ebers-Moll model provides an adequate description in certain ranges of current and voltage, which traditional theory suggests are determinable [33] in terms of the onset of failure of these approximations, and for frequencies below certain critical frequencies set by the one-pole approximations of the current gains. As will be seen, the other existing models do not depend upon all of these approximations and consequently have wider ranges of validity.

2.4-3 Modified Ebers-Moll Models for CIRCUS, SCEPTRE, NET, CANCER, SLIC, etc.

The models associated with many computer circuit analysis programs--including CIRCUS, SCEPTRE, NET, CANCER, SLIC--all are straightforward generalizations of the original Ebers-Moll model. In these generalizations, one treats certain parameters that are constants in the Ebers-Moll formulation as functions of the transistor currents and voltages. This scheme retains the form of the original Ebers-Moll model but widens the range of currents and voltages over which an adequate description prevails.

Many different ways exist to specify the functional dependence of the parameters. Commonly the current gains (β_F and β_R; α_F and α_R) and the time constants (τ_{11} and τ_{22}) are specified as functions of the collector current. Each functional dependence is determined by measurements made at the device terminals and each may be tabulated for inclusion in the model description. Alternatively, one can include a dependence by fitting data to a polynomial, as is done, for example, in the model associated with the NET II program [14]. Or certain functional dependences may be included by adding an electric circuit element, such as a resistor, to the model; the value of the element again derives from fitting data provided by terminal measurements. In SCEPTRE, one uses this approach in representing the cut-off model of transistor operation [13]. In CANCER [38] and SLIC [15], it is used to describe the Early effect produced by base-width modulation [34]. Logan [4,5,39] and Lindholm,et al. [40,41] have proposed a similar alteration to account for the Early effect. It consists in applying measured data to specify a new parameter, a critical voltage, which is used to form a factor that multiplies collector current and current gain. From the device terminals, the result is like that produced by adding a nonlinear resistor appropriately in the model.

Fitting data by addition of a circuit element may ease computational problems. This advantage aside, the approach differs basically not at all from the other curve-fitting schemes.

The functional dependence of the slope factors (the θ's in Eq. 9) may also be determined from measured data, and this is done in some modified Ebers-Moll models. Because the slope factors appear in exponents in the model formulation, predicted device behavior depends with immense sensitivity on the functional dependence given these factors.

The strength of these curve-fitting models is accuracy. Indeed, to gain more accuracy, one presumably needs only to incorporate more measured data into the model description. Because curve fitting ignores the internal mechanisms that yield observed behavior, it can give accurate predictions of device performance irrespective of which internal mechanisms predominate. In principle, therefore, the class of modified Ebers-Moll models can account for high injection, breakdown, multi-dimensional flow or whatever other phenomena are present that invalidate the original Ebers-Moll model. In practice, however, many of the existing versions of the modified Ebers-Moll model fail to embody enough data in a sufficiently clever way to simulate device performance in the presence of certain internal mechanisms. For example, many make no attempt to describe behavior accompanying breakdown. Evidently none can accurately predict transistor performance in the saturated or inverse-active modes of operation, where carriers injected from a large collector may flow along severely multi-dimensional paths to be collected at a small emitter. But present shortcomings such as this do not appear to be fundamental. If the goal is to describe the behavior of a single discrete transistor, nothing basic appears to stand in the way of securing more and more accuracy, whatever the mode of behavior and whatever internal mechanisms prevail, by incorporating into the model more and more data measured at the terminals.

While the strength of a modified Ebers-Moll model derives from ignoring internal mechanisms, so also do its weaknesses. In curve-fitting schemes, one gains accuracy at the cost of great complexity. But model complexity may limit the size of circuit amenable to computer simulation, an important issue as we enter an era of MSI and LSI circuits. This consideration--together with considerations of the time, cost, and difficulty involved in accurately measuring the functional dependence of the model parameters--may make questionable the economics of doing a simulation at all. Furthermore, only because these curve-fitting models are based in the physics underlying the original Ebers-Moll model is there any direct link to the processing used in making the transistor and hence to transistor design. By ignoring such internal mechanisms as high injection and crowding, these models give no systematic guidance for designing transistors to operate in circuits that cause voltage-

current excursions outside the range of validity [33,42] of the original Ebers-Moll model.

There are many different variations of modified Ebers-Moll models, and many more possible than have yet appeared. Each derives, however, from the original Ebers-Moll model or its equivalents. For example, the intrinsic models associated with SCEPTRE, NET and CANCER derive from the injection form of the Ebers-Moll model or its equivalents. For example, the intrinsic models associated with SCEPTRE, NET and CANCER derive from the injection form of the Ebers-Moll model, though they employ some notation related to the charge-control model. The CIRCUS model derives from the transport form of the Ebers-Moll model; it likewise uses some charge-control notation. The large-signal model associated with SLIC has its origin in the transport form rearranged to yield a topology similar to that of Howard's nonlinear hybrid-pi model [43] (which mathematically is equivalent to the original Ebers-Moll model).

Their origin in the Ebers-Moll formulation implies that all these models, and indeed the whole class of modified Ebers-Moll models, can be represented by the general circuit diagram of Figure 5. Each is a special case. Eqs. (28) through (37) still describe the elements of this circuit diagram, provided one understands that the functional dependence of the β's, τ's and θ's are to be included by tabulation or specified implicitly by one of the methods discussed earlier in this section.

2.4-4 Fossum's Expandable Model

Fossum [19,20] has suggested a different generalization of the Ebers-Moll model. This generalization exploits the advances in device theory made since the appearance of the Ebers-Moll paper in 1954. During these nearly two decades, the processing by which transistors are fabricated has changed profoundly, and consequently so too have the details of transistor behavior. In response to these changes, device theorists have proposed explanations attributing different aspects of the observed behavior to the dominance of different sets of physical processes occurring within the transistor materials. Fossum proposes a transistor model containing the currently accepted theoretical representations of each of these processes.

To describe his approach, Fossum suggests the term, expandable Ebers-Moll modeling. The modeling remains based in the original Ebers-

Moll formulation but is expandable in the sense that it includes processes--base-width modulation and high injection, for example--ignored in the original Ebers-Moll model. Its basis in device theory distinguishes expandable modeling from the modeling described in Section 2.4-3 associated with the various computer circuit analysis codes. In an expandable Ebers-Moll model, one accounts for such observed behavior as the dependence on current of transistor current gain by representing fundamental physical processes individually, modeling in closed functional form. The computer analysis program being used then automatically superimposes the effects of the various processes, yielding a functional dependence on the terminal variables of the current gain and other relevant measures of device performance. In modified Ebers-Moll modeling, one accounts for such dependence either by tabulation of transistor performance measured at the device terminals or by some of the other methods discussed in Section 2.4-3, which are in essence equivalent to tabulation.

Because of its basis in device theory, expandable modeling couples more closely with device and integrated-circuit fabrication and in this sense more closely with design. At present, however, insufficient evidence exists to compare, for different types of transistors in different circuit applications, the relative merits of the two modeling techniques as tools for semiconductor circuit analysis. Questions of comparative accuracy, computer time, and memory requirements remain unanswered.

Figure 9 illustrates an expandable model for an npn transistor. In this model each element describes a different physical process: i_{EN} and i_{CN} represent current flowing in response to minority carrier injection into the base region from the emitter and the collector; i_{ER} and i_{CR} represent current flowing in response to recombination and generation of free carriers in the junction space-charge regions; $\alpha_{T1}i_{CN}$ and $\alpha_{TN}i_{EN}$ represent the portion of the current emitted from one junction that is transported across the base region to be collected by the other junction; i_{EP} and i_{CP} represent the current resulting from injection of minority carriers from the base region into the emitter and collector regions; the four current sources involving the multiplication factors M_E and M_C describe the results of avalanche multiplication and breakdown

at the junctions. In addition to the processes represented by these twelve current sources, the model contains four capacitors. Junction depletion capacitors are labeled C_{JE} and C_{JC}. Through these capacitors flow displacement currents that change the charge associated with the junction space-charge regions. The large-signal counterparts of the diffusions capacitors are labeled C_{DEN} and C_{DCP}. Through these capacitors flow displacement currents that change the excess carrier concentrations stored in the quasi-neutral emitter base and collector regions.

In the model, the effects of high injection and base width modulation are accounted for. To maintain one-to-one correspondence with physical processes, these effects appear represented in the theoretical descriptions of all sixteen of the model elements. The effects of net recombination in the junction regions are accounted for by the sources i_{ER} and i_{CR}; recombination at the surface can be represented by similar sources, not shown in Figure 9. Avalanche breakdown is taken into account by the four current sources involving M_E and M_C. As we noted earlier, the original Ebers-Moll model neglects high injection, base-width modulation, and net recombination in the junction regions and at the surface. Because the expandable Ebers-Moll model includes these phenomena, it provides more accuracy over a wider range of currents and voltage.

The present version of the expandable Ebers-Moll model fails, however, to include the effects of other phenomena that can influence transistor behavior. Chief among these are the effects of multi-dimensional flow and of high doping. These shortcomings are present in all existing models, and we discuss them in the final section of this report.

As was pointed out in Section 2.4-3 for the modified Ebers-Moll models, however, present shortcomings should not be construed to be ineradicable deficiencies. In this spirit, as Fossum suggests, it is instructive to distinguish between any one expandable model and the general technique of expandable modeling.

Figure 9, representing the expandable Ebers-Moll model, has the same topology as the circuit diagram of Figure 6. From Section 2.4-1, therefore, it follows at once that the expandable model is a special case of the circuit diagram of Figure 5. Moreover, as might be

expected because of its origin in the original Ebers-Moll model, currents i_{EN} and i_{CN} in the expandable model have an exponential dependence that make it correspond to the matrix description formed by combining Eqs. (9) and (13). In contrast with the original Ebers-Moll formulation, however, the slope factors θ_{ET}, θ_{CT}, θ_{ER}, and θ_{CR} appearing in Eq. (9) are unequal in the expandable model. Fossum chooses

$$\theta_{ET} = \theta_{CT} = \frac{e}{kT} = 2\theta_{ER} = 2\theta_{CR} \tag{38}$$

in accord with present transistor theory.

As before, to specify for the expandable model the functional dependences of the element values in Figure 5 requires that first we identify the leakage currents i_{Y1} and i_{Y2} and then substitute into Eqs. (20 and (21). From examining the base current i'_B flowing in the model of Figure 9 for static excitation we conclude that

$$i_{Y1} + i_{Y2} = (1- \alpha_{TN}M_C)i_{EN} + i_{EP} + M_E i_{ER} + (1-\alpha_{TI}M_E)i_{CN} + i_{CP} + M_C i_{CR} \tag{39}$$

Because terms like $\alpha_{TN}M_C i_{EN}$ depend both on v'_{BE} and v'_{BC} in the expandable model, the choice for i_{Y1} and i_{Y2} is less clear than it was for the original Ebers-Moll model. We choose

$$i_{Y1} = (1-\alpha_{TN}M_C)i_{EN} + i_{EP} + M_E i_{ER}$$
$$i_{Y2} = (1-\alpha_{TI}M_E)i_{CN} + i_{CP} + M_C i_{CR} \tag{40}$$

Other choices are possible. This one has the merit of being consistent with choices made before, in Eqs. (26) and (27). Moreover, combined with Eq. (20), it leads to an expression for the transport current,

$$i_X = \alpha_{TN}M_C i_{EN} - \alpha_{TI}M_E i_{CN} , \tag{41}$$

which reduces to Eq. (27) in the absence of carrier multiplication ($M_C = M_E = 1$). From Eq. (21)

$$C_1 = C_{JE} + C_{DEN} \; ; \; C_2 = C_{JC} + C_{DCP} \tag{42}$$

For the functional dependences to be assigned to the various parameters appearing in Eqs. (40) through (42), one should consult Fossum [19,20].

2.4-5 The Gummel-Poon Integral Charge Control Model

Of all the models for the intrinsic bipolar transistor, the integral charge-control model relates most directly to the circuit representation of Figure 5. Although Gummel and Poon never make explicit this relation, it becomes plain from examining their derivation.

The derivation of the integral charge-control model starts from the transport form of the Ebers-Moll model, expressed in Eq. (23). Written out for dc excitation, this equation becomes:

$$I_C = I_{CC} - I_{B2} \tag{43}$$

$$I_E = -I_{CC} - I_{BJ} \tag{44}$$

$$I_B = -(I_E + I_C) = I_{B1} + I_{B2} \tag{45}$$

Gummel and Poon term

$$I_{CC} = I_S[\exp(eV'_{BE}/kT) - \exp(eV'_{BC}/kT)] \tag{46}$$

to be the dominant component of the collector current, and

$$I_{B1} = (I_S/\beta_F)[\exp(eV'_{EB}/kT)-1] \tag{47}$$

and

$$I_{B2} = (I_S/\beta_R)[\exp(eV'_{BC}/kT)-1] \tag{48}$$

constitute the components of the base current. In the integral-charge-control model, one retains the form of Eqs. (43) through (45) but gives new characterizations for the dominant component of the collector current and for the components of the base current. These new characterizations extend the range of currents and voltages over which the model adequately describes transistor performance, widening it beyond that provided by the original Ebers-Moll formulation.

Before describing these characterizations, we digress briefly to point out the relation to the circuit representation of Figure 5. The dominant component of the collector current corresponds to the current source i_X, defined earlier, representing transport. The components of the base current correspond to the current sources i_{Y1} and i_{Y2}, defined earlier, representing leakage. Thus, for dc excitation, the separation into the components indicated in Eqs. (43) through (45) implies at once representation by the circuit of Figure 3. For time-

varying excitation, use of the quasi-static approximation described in Section 2.3 then implies representation of the integral-charge-control model by the circuit of Figure 5. The work of Chawla [17] makes this equivalence evident from another viewpoint, as we shall soon see.

The Gummel-Poon characterization of the dominant component of the collector current can be derived from the basic equations of semiconductor device physics. As shown by Gummel [44], the dominant component can be expressed as

$$i_X = i_{CC} = (Q_{BO}/Q_B)I_S[\exp(ev'_{BE}/kT) - \exp(ev'_{BC}/kT)] \quad (49)$$

where

$$Q_{BO}I_S = I_k^2\tau_f \exp(-v_k) \quad (50)$$

in terms of the parameters I_k, τ_f, and v_k measurable at the device terminals [18]. Eq. (49) differs from the Ebers-Moll expression of Eq. (46) only by the multiplicative factor Q_{BO}/Q_B. Here Q_B represents all charge in the transistor residing on carriers of the type that communicate with the base terminal. For an npn transistor, this is the charge on all of the holes in the transistor. In thermal equilibrium with every source of applied excitation absent, Q_B by definition equals Q_{BO}. For such conditions that Q_B nearly equals Q_{BO}--the condition of low injection, for example--the Gummel expression of Eq. (49) reduces to the Ebers-Moll expression of Eq. (46). For more general conditions, however, the Gummel expression enables inclusion of effects ignored in the Ebers-Moll model.

The extension given by Eq. (49) is one of the key facets of the integral charge control model, enabling prediction of transistor performance for currents and voltages outside the range of validity of the original Ebers-Moll model. To understand the physical basis for the generalization contained in Eq. (49), it is instructive to contrast the assumptions underlying the integral charge-control and the Ebers-Moll model. A main distinction between these two models arises from the charge focussed upon. Ebers and Moll chose to focus attention on the charge on the minority carriers in the base region, which are electrons for an npn transistor. Gummel and Poon center attention on the charge on all of the holes throughout the whole of an npn transistor. For the transistors available to Ebers and Moll in 1954, the wide base

regions justified their choice. Transistor behavior was dominated by the behavior of minority carriers in the quasi-neutral base. For a transistor of the present day, however, the minute size of the base region makes significant the role played by free carriers located outside the quasi-neutral base. The choice of Gummel and Poon is thus the more reasonable one for modern transistors.

Utilization of Gummel's charge-control expression of Eq. (46) requires that one specify the dependence on terminal currents and voltage of the controlling charge Q_B. This specification is the second key facet of the integral charge-control model. Through the functional dependence given Q_B, one can model such effects as high injection in the base region (conductivity modulation) [45], base push-out [46], and base-width modulation [34] or the Early effect--all of which the original Ebers-Moll model ignores. There are many choices possible for this functional dependence. One that has proved useful is [16,18]

$$Q_B = Q_{BO} + Q_E + Q_C + B\tau_f i_F + r_t \tau_f i_R \tag{51}$$

Here τ_f is the forward delay time (or charge-control constant) - a parameter which can be measured at the device terminals [18]. The parameter r_t is the ratio of reverse-to-forward delay time and is likewise determinable [18].

In terms of measurable parameters [18] the remaining quantities in Eq. (50) are as follows.

$$Q_{BO} = -I_k \tau_f \tag{52}$$

Q_E/Q_{BO} = charge associated with emitter junction capacitance

$$= R(ev'_{BE}/kT, P_e) \tag{53}$$

Q_C/Q_{BO} = charge associated with collector junction capacitance

$$= R(ev'_{BC}/kT, P_c) \tag{54}$$

where

$$R(v,P) = p_3 \left\{ \frac{1}{(1+p_4)^{p_2}} + \frac{(v/p_1 - 1)}{[(v/p_1-1)^2 + p_4]^{p_2}} \right\} \tag{55}$$

and P_e and P_c are parameters relating to the emitter and collector transistion capacitances [18]

$$B = \text{base-push-out factor} = 1 + r_w \frac{\left[\sqrt{i_4^2+r_{p1}}-i_4\right]^{n_p}}{4(I_C^2/I_k^2+r_{p2})} \qquad (56)$$

where

$$i_4 = (i_C/I_k) + \frac{v_{OC}-ev'_{BC}/kT}{v_{rp}} . \qquad (57)$$

Finally,

$$i_F = (I_k^2\tau_f/Q_B)\exp(-v_k)\exp(ev'_{BE}/kT)$$

$$i_R = (I_k^2\tau_f/Q_B)\exp(-v_k)\exp(ev'_{BC}/kT) \qquad (58)$$

The meaning of each quantity undefined above appears in Table 1 of Poon [18]. They all are measurable parameters.

The expression for the transport current given in Eq. (49) is grounded in rigorous semiconductor device theory. For the most part, so also are the allied expressions that described the controlling charge Q_B. In contrast, Gummel and Poon rely on semi-empirical descriptions of the leakage current, forced to this reliance by lack of detailed knowledge of the properties and spatial distribution of the defect centers. If avalanche effects are included according to the theory of Poon and Meckwood [47], these descriptions of the leakage currents become easily identifiable as

$$i_{Y1} = -I_k\{i_1\exp(-v_k)[\exp(ev'_{BE}/kT)-1] + i_2\exp(-v_k/n_e)[\exp(ev'_{BE}/n_ekT)-1]\} \qquad (59)$$

$$i_{Y2} = -I_k i_3\exp(-v_k/n_c)[\exp ev'_{BC}/kT-1] - i_A \qquad (60)$$

in which

$$i_A = \left(\frac{2\alpha_N}{b_n}\right) i_C \left|(kTv_{OC}/e) + v'_{BC}\right| \left[1 + \frac{\eta E_o}{\sqrt{(kTv_{OC}/e)+v'_{BC}}}\right] \exp(-b_n/E_m) \qquad (61)$$

is the component of current arising from avalanche multiplication in the collector space-charge region [47]. In these expressions, α_n is the avalanche coefficient, b_n is the critical field for avalanche, and the remaining undefined quantities are determinable model parameters [18].

As we have seen, Gummel and Poon's derivation of their model makes clear that it can be represented by the circuit of Figure 5. This can also be demonstrated from the work of Chawla [17]. Chawla made explicit the dynamic properties of the integral charge control model, formulated for dc static excitation by Gummel and Poon, and demonstrated a circuit representation of the model having the topology of Figure 6. From this it follows at once that the integral charge control model is a special case of the circuit of Figure 5.

The current sources in this circuit have the functional dependences given in Eqs. (49), (59), and (60). The capacitors are described by [17]

$$C_1 = \frac{\partial Q_B}{\partial v'_{BE}} = \frac{\partial (B\tau_f i_F)}{\partial v'_{BE}} \tag{62}$$

$$C_2 = \frac{\partial Q_B}{\partial v'_{BC}} = \frac{\partial (r_t \tau_f i_R)}{\partial v'_{BC}} \tag{63}$$

as becomes clear from consideration of the quasi-static approximation used in deriving time-varying behavior. Eqs. (53) through (58) contain the functional dependence of charges $B\tau_f i_F$ and $r_t \tau_f i_R$ and hence of capacitors C_1 and C_2.

In the absence of avalanche effects, the equations describing the integral charge control model fit directly as special cases of the matrix formulation composed of Eqs. (9) and (13). The pertinent slope factors are implied in the Gummel-Poon constants n_e and n_c. If avalanche effects are included, however, the fit is more strained. To accommodate the avalanche current i_A described by Eq. (61) one can choose an element of the [R] matrix appropriately and set the corresponding slope factor to zero. Less obliquely, one can add to Eq. (9) a current matrix having the dependence needed to include the avalanche current.

If the controlling charge Q_B is characterized in the full detail implied in Eqs. (51) through (58), the integral charge control model includes a characterization of the effects of base-width modulation, high injection, and base push-out. If the base current is characterized in the full detail implied in Eqs. (59) through (61), the integral charge control model includes a characterization of the effects of recombination generation and avalanche multiplication in the junction space-charge regions. This characterization of the base current, and to a lesser

degree of the controlling charge, enables prediction of how the current gain and other measures of device performance depend on the terminal currents and voltages. The integral charge control model is based in semiconductor device theory. Thus for the most part, its parameters relate to the parameters of transistor fabrication; the model is meant to serve as a guide in device design as well as a tool for the simulation of circuit behavior.

The integral charge control model, however, like the other models we have discussed, ignores the presence of multi-dimensional flow in its derivation. Such effects as current crowding [48-50] or lateral spread attending high current in the collector [46,51] or multi-dimensional flow present for operation in the saturation or inverse-active models all are ignored in the derivation. They can be included only by fitting the values of certain parameters to match data measured from the terminals; Gummel and Poon suggest methods for doing this [16]. These methods give information oblique to device fabrication and design and hence oblique to *a priori* design by computer. They may accomplish the inclusion of multi-dimensional flow in the simulation of discrete circuit behavior, however, just as, in principle, the modified Ebers-Moll modeling of Section 2.4-3 can include it. How accurately multi-dimensional effects are included remains a question. The answer awaits the results of future experimental and computational studies for a wide range of different transistor types in a wide range of different circuit applications.

As to the other existing models for circuit analysis, the integral charge control model neglects the effects of heavy doping. These we discuss in Section 2.8.

2.5 Modeling the Extrinsic Device

Until now our attention has centered on the intrinsic part of the transistor, enclosed by the dotted lines shown in Fig. 2(a). To a first approximation, the densities of mobile carriers in the extrinsic part of the transistor rest at their thermal equilibrium values; with this approximation made, the extrinsic part is modeled more easily than is the intrinsic part. If dc steady-state conditions prevail,

for example, the predominate mechanisms occurring in the extrinsic part is current conduction to the base, collector, and emitter terminals; and this conduction is modeled by resistors. For general time-varying conditions, however, the model becomes an RC transmission line, as illustrated in Fig. 10(a). The parameters of the line are nonlinear to the extent that the applied excitation at the terminals influences the height of each potential barrier at pn junctions located in the extrinsic part.

Because partial differential equations describe RC transmission lines, in the complex frequency domain a matrix formulation of their behavior involves transcendental functions of the frequency variable s. To secure a model tractable for circuit analysis, computer-aided or not, one commonly approximates each transcendental function by polynomials or ratios of polynomials in s. The approximations are similar to those used for the [T] matrix of Eq. (9). Many ways exist to make lumped-circuit approximations of this kind [2]. In Fig. 10(b), we show a simple example. Fig. 10(c) shows the resulting overall model formed by combining with the model of Fig. 5(b) for the intrinsic device.

To provide a first-order accounting of the effects of multi-dimensional flow occurring in the inverse-active and saturation regions of operation, one can represent the collector-base junction in the extrinsic part by a pn junction diode connected between collector and base. This is often called an overlap diode [26].

A fully accurate representation of high-frequency or fast-transient behavior requires more complicated models for the extrinsic part than those in Fig. 10(b) [2,52-55]. For microwave applications [52-55] the requirements may become especially severe. Models no more complicated in topology than that of Fig. 10(b) - and containing nothing more than resistors, capacitors, and dependent sources - are now widely used, however, for computer simulation in digital electronics.

2.6 Generalization of Results: IGFET or MOS Transistor

Though to this point our attention has fixed upon the bipolar transistor, many of the results and conclusions of Sections 2.1 through 2.3 hold also for the insulated-gate field-effect transistor (IGFET) or

MOS transistor, and for field-effect transistors of other types: the junction field-effect transistor (JFET) and the Schottky-barrier or metal-gate field-effect transistor (MESFET). Here we point out similarities and differences.

The structure of an IGFET causes electric field, current and carrier density, and other pertinent variables to depend upon three spatial corrdinates. Figure 11, displaying this, is the counterpart of Figure 1 for the bipolar transistor, which showed that the bipolar structure gives rise to three-dimensional dependence. As is common practice for the bipolar transistor, one evades the resulting three-dimensional boundary-value problem by ignoring dependence in the z coordinate direction, thus reducing the problem to one involving only two dimensions. The two-dimensional model for the IGFET appears in Fig. 11(b), and is parallel entirely to Fig. 1(b) for the bipolar transistor.

For this two-dimensional IGFET structure, the equations of semiconductor device physics [21,22], combined with Maxwell's equations have been solved rigorously by computer, with main attention paid to the dc steady state [56-59]. The numerical results of this computer solution prove intractable for direct use in the analysis of circuits containing many devices, though they have proved useful [56] in the derivation of an analytical expression for drain current suitable for incorporation in circuit models. Existing circuit models, however, all derive from a reduction-of-dimension approximation. This approximation [60] reduces the two-dimensional boundary-value problem to coupled one-dimensional problems, which describe the intrinsic part, combined with side conditions, which describe the extrinsic part. Figure 12 shows the sectioning into intrinsic and extrinsic parts.

In general direction, these procedures follow exactly those employed in characterizing the bipolar transistor, illustrated in Figure 3. Additionally, the treatment of the intrinsic part of the IGFET exactly parallels that displayed in Figures 4 and 5. The resulting circuit model is shown in Figure 13, in which the notation underscores that this model describes equally the MOS and the bipolar transistor.

The circuit model of Figure 13, combined with appropriate models for the extrinsic part [60], contains as special cases all of the main models proposed earlier. In an earlier publication [60], the author has demonstrated the detailed connection to several models previously proposed for the computer-assisted analysis of MOS circuits. Additionally, MOS models associated with circuit analysis codes such as SCEPTRE [13], the model proposed by Meyer [61], and the model of Jenkins et al. [62] are represented by the circuit diagram of Figure 13, if appropriate elements are added to account for the extrinsic part [60].

Figure 3 applies to IGFET's equally well as it does to bipolar transistors. For both devices, very similar approximations lead to a equivalent circuit of the same topology, which represents the intrinsic part of either the bipolar or MOS transistor. Thus the same matrix, given in Eq. (13), describes either device.

The models for the two kinds of devices differ, however, in that the functional dependence of the current sources and the capacitors may profoundly differ. For example, the current source i_x, representing transport, shows for the bipolar transistor an exponential dependence upon the terminal voltages, contrasting with the power law dependence pertaining to the MOS transistor (operated in strong inversion). Differences exist also in the representations used to model the extrinsic part [60].

2.7 Phenomenological Inclusion of Radiation Effects in Bipolar Transistor Models

We shall treat two kinds of radiation effects in bipolar transistors:

(a) radiation-induced changes in the microstructure of the semiconductor--often termed permanent radiation damage; and

(b) radiation-induced photo-currents--often termed transient or pulse effects.

Thermal, mechanical or macro-structural changes may also result, but these lie beyond the scope of this discussion.

From a phenomenological viewpoint, both kinds of radiation effects listed above enter any of the transistor circuit models in the same way. No new circuit elements need be added to the model of Figure 5 for the intrinsic transistor because the same general kind of phenomena persist in the absence or presence of radiation effects. These phenomena are

transport processes, generation-recombination-trapping-tunneling processes (termed leakage processes in Figure 5), and charge-storage processes. Radiation can drastically alter the details of the physics needed to describe these processes, however, and the functional dependence of each circuit element in the model must take these alterations into account.

In the sense that each step used in fabricating a transistor changes the microstructure of the device materials, one can regard permanent radiation damage as an additional fabrication step. As one of its main results, this step produces flaws within the crystalline structure of the silicon. Depending on the location and distribution of the additional allowed energies associated with these flaws and the quantum-mechanical transition probabilities among allowed energies in the energy-band structure, the flaws will likely act in one of two ways. Either they act as recombination-generation centers, increasing the rate of transitions between the conduction and the valence bands. Or they act as electron or hole traps, removing majority carriers and increasing the resistivity of the material.

Other possibilities also exist. Flaws located in junction regions of heavy doping could increase the probability of quantum-mechanical tunneling or field emission, thus limiting the range of usable terminal voltages that can be applied. Or flaws could increase either recombination generation or trapping through Auger processes, which may accompany the high carrier densities attending radiation-induced photocurrents. An instructive tabulation of the possible transitions associated with flaws in semiconductor material appears in a paper by Sah [63].

All of these possibilities are included, without need for additional circuit elements, by the phenomenological circuit representation of Figure 5 or its equivalent, Figure 6. More detailed equivalent circuits are available, due to Sah [64-67], and Raymond [68,69]. Through multi-section representation of the internal processes, these equivalent circuits better describe the distributed nature of device materials. Moreover, they may lend insight about the details of the physics underlying device operation and give greater accuracy in the description of device behavior. But as yet these more detailed representations have found much less usage in circuit-analysis codes than the models represented in Figures 5 and 6.

Though permanent radiation damage leaves unchanged the circuit topology of the model for the intrinsic transistor, it will alter the parameter values of each circuit element in the model. For example, consider the integral charge-control model. Removal of majority carriers produced by radiation damage will reduce Q_{BO}, causing the shift in current-voltage characteristics implied in Eq. (49). If radiation introduces a substantial number of flaws, the slope factors n_e and n_c can change, which, as Eqs. (59) and (60) imply, can likewise produce sizeable changes in the characteristics. These occurrences, and others not mentioned here in this brief discussion, have their counterparts in each of the other models we have discussed. Radiation also can give rise to significant changes in the values of the circuit elements of the model for the extrinsic transistor and in the functional dependences exhibited by these circuit elements.

As was noted earlier, from a phenomenological viewpoint both permanent and transient radiation effects are entered into the intrinsic model without addition or deletion of circuit elements. Transient radiation produces photocurrents. The current generators i_{Y1} and i_{Y2} account for photocurrents, treating them as a special type of generation-recombination process.

For both permanent radiation effects and transient effects, therefore, the issue is to specify the pertinent functional dependences of each circuit element in the model of Fig. 5 for the intrinsic transistor and in the model for the extrinsic transistor. One can do this either from considerations of physics or from inferences drawn from electrical measurements made at device terminals. Much effort has been devoted to these problems [70]. But it is beyond the scope of this report and the expertise of the author to assess the relative value of the choices now available.

2.8 Discussion

We have shown that most of the large-signal circuit models in common use for the intrinsic bipolar transistor, and for the intrinsic field-effect transistor, can be represented by the circuit diagram of Fig. 5. The differences among these models lies in the functional dependence assigned to each of the two capacitors and three current sources in this circuit diagram.

Two methods exist for determining this functional dependence:

(a) by inferences made from measurements at the device terminals; and

(b) by drawing from the physics-based understanding provided by semiconductor device theory.

Certain of the models emphasize one method as opposed to the other. Certain others admit use of a combination of the two methods.

The set of modified Ebers-Moll models, which are the models most widely used in present-day computer-aided circuit analysis, rely strongly on use of measurements at the device terminals. In current practice, the Gummel-Poon integral charge-control model likewise emphasizes use of electrical measurements for the determination of its parameters, even though this model is linked more closely to the underlying physics than is the set of Ebers-Moll models. Indeed the most sophisticated facility known to the author for the automatic determination of parameters has been based on the integral charge control model [71].

Of the existing circuit models discussed in this report, Fossum's expandable model places the most reliance on semiconductor device theory for the determination of its parameters. By making his model expandable, in the sense described earlier, Fossum recognizes that limitations of the existing semiconductor device theory prevent a fully accurate description of the behavior of modern bipolar transistors.

In the author's opinion, the following constitute some of the major sources of these limitations:

(a) inadequate inclusion of multi-dimensional effects;

(b) inadequate inclusion of the effects of high doping;

(c) inability to calculate doping profiles accurately; and

(d) reliance on physics questionable for small devices.

A full discussion of these considerations is beyond the purpose of this report, but a brief indication of each can be given.

Multi-dimensional flow occurs for several reasons. In saturation or in the inverse-active mode of operation, it may arise from the difference in area of the collector and emitter regions. In the forward-active region, device theory, as presently advanced, associates multi-dimensional flow with emitter crowding [48-50] and with modes

possible when high current flows in a lightly-doped collector region [46,51]. Other origins of multi-dimensional flow may also exist [23].

The modified Ebers-Moll models can take into account the effect of multidimensional flow. But, as described earlier, they do so via curve-fitting derived from measurements at the device terminals. Of the physics-based models, the present version of the expandable model takes no fundamental account of any of these effects, though inclusion of the extrinsic base resistance may suggest zeroth order trends of the effects of crowding. Neither does this version include base push-out [46], which existing device theory considers a one-dimensional phenomena [46], other than the zeroth order trends suggested by the presence of the extrinsic collector resistance. The integral charge-control model does a zeroth-order modeling of multi-dimensional effects by allowing the value of one of its parameters to fit data measured at the device terminals.

High doping concentrations in silicon [72,73] have been shown to play a first-order role in certain transistors in determining the current gain and the speed [74-77]. None of the physics-based circuit models discussed in this report include the effects of high doping. The expandable model of Fossum is designed, however, to accommodate new findings in semiconductor device theory, and the effects of high doping can be imbedded in an updated version of this model.

The theory of the effects of high doping is presently in flux [77]. From researchers in Belgium has come the computer program SITCAP [78], which contains one version of the theory of these effects. This program includes also a first-order modeling of multi-dimensional effects, following the method of approximation first suggested by Ghosh [49]. The intent of SITCAP is not to serve directly as an equivalent circuit for the analysis of sizeable transistor circuits. Rather, it seeks to link impurity profiles to dominant physical processes to electrical behavior at the terminals.

To secure circuit models needed for the complete *a priori* design by computer of an integrated circuit, an additional link is needed: a link between the impurity profiles and the parameters of manufacturing. At present, one is unable to calculate accurately the impurity profiles constituting a transistor from knowledge of surface concentrations, diffusion times, etc.

The small size of existing devices, and the trend toward ever smaller size, raises questions about the appropriateness of the basic assumptions upon which all analysis of transistors is based. As one example, this analysis relies on the assumption that current flows by drift and diffusion. In small high-speed transistors, however, the quasi-neutral base width may be so small that carriers crossing it fail to experience many collisions. This places into question the appropriateness of drift and diffusion as the mechanism of carrier flow. None of the existing circuit models can include the effects consequent to this possible misuse, which only recently has been suggested [79].

Having this perspective of the limitations of the existing circuit models for the bipolar transistor, we can make some practical comments about the design of a system using bipolar integrated circuits. If the system can be designed with standard products (shift registers, adders, etc.), black box modeling [80] may black box modeling ignores totally the physics underlying the operation of an integrated circuit suffice. If, however, the system requires custom integrated circuits, manufactured in small numbers, fundamental reasons [81] necessitate that the circuit models be compatible with *a priori* design of the integrated circuits by computer. These circuit models must be compact enough mathematically to enable fast analysis of circuit behavior, and their parameters must be accurately calculable from knowledge of pertinent steps in manufacturing.

None of the existing circuit models for the bipolar transistor meets these requirements. Though they are presently the models most widely used in circuit analysis, the set of modified Ebers-Moll models is unsuited to a complete *a priori* design by computer because they rely on curve fitting to measurements made at the transistor terminals. In the opinion of the author, their widespread use has arisen as an engineering necessity forced by the inadequacies of the existing theory of semiconductor devices. The models of Fossum and of Gummel and Poon, which are more soundly based in device physics, suffer from these same inadequacies.

As a final conclusion, therefore, the author believes that circuit models for bipolar transistors is far from a finished subject, despite

the years that have been invested in their development. Very practical considerations require circuit models compatable with the *a priori* design of integrated circuits by computer. In turn, this requirement raises the need for improvements in semiconductor device theory designed to support the development of such circuit models.

References

1. D. J. Hamilton, F. A. Lindholm, and J. A. Narad, Proc. IEEE, 52, 248 (1964).

2. D. J. Hamilton, F. A. Lindholm, and A. H. Marshall, Principles and Applications of Semiconductor Device Modeling, Host, Rinehart and Winston (1971), chapter 11.

3. J. Lange, IEEE Trans. on Elec. Dev., Dec. 1971.

4. J. Logan, Bell System Tech. Journal, 50, 1105-1147 (April 1971).

5. J. Logan, Proc. IEEE, pp. 78-85 (Jan. 1972).

6. G. Aaronson and R. B. Schilling, IEEE Trans. on Elec. Dev., pp. 1-3 (Jan. 1972).

7. J. J. Ebers and J. L. Moll, Proc. IRE. 42, 1761 (1954).

8. J. L. Moll, Proc. IRE. 42, 1773 (1954).

9. R. Beaufoy and J. J. Sparkes, A.T.E. Jl. 13, 310 (1957).

10. C. Rosenberg, D. S. Gege, R. S. Caldwell, G. H. Hanson, IEEE Trans. Nuc. Sc., NS-10, 149 (1963).

11. J. G. Linvill, Proc. IRE. 46, 949 (1958).

12. L. D. Miliman, W. A. Massena and R. H. Dickhaut, "CIRCUS, Digital Computer Program for Transient Analysis of Electronic Circuits -- User's Guide," Harry Diamond Lab. Tech. Rept. 346-1, Jan. 1967.

13. J. G. Bowers and S. R. Sedore, SCEPTRE: A Computer Program for Circuits and Systems Analysis, Englewood Cliffs, N. J.: Prentice Hill, 1971.

14. A. F. Malmberg, "NET-2 Network Analysis Program -- Preliminary Users's Manual," Harry Diamond Lab., May 1970.

15. T. E. Idleman, F. F. Jenkins, W. J. McCalla and D. O. Peterson, IEEE Journ. Solid-State Circuits, SC-6, 188 (1971).

16. H. K. Gummel and H. C. Poon, Bell Syst. Tech. Journal, 49, 829 (1970).

17. B. R. Chawla, IEEE Journ. Solid-State Circuits, SC-6, 262 (1971).

18. H. C. Poon, IEEE Trans. Electron Devices, ED-19, 719 (1972).

19. J. G. Fossum, Proc. IEEE, 60, 756 (1972).

20. J. G. Fossum, IEEE Trans. Electron Devices, ED-19, 582 (1973).

21. W. Shockley, B.S.T.J., 28, 435 (1949).

22. W.Van Rossbroeck, B.S.T.J., 29, 560 (1950).

23. D. P. Kennedy, Final Report: AFCRL-72-0257, January 31, 1972.

24. J. W. Slotboom, IEEE Trans. Electron Dev., (to appear).

25. M. S. Ghausi and J. J. Kelly, Introduction to Distributed-Parameter Networks, Holt, Rinehart and Winston, 1968.

26. C. S. Meyer, D. K. Lynn, and D. J. Hamilton, Analysis and Design of Integrated Cirtuits, McGraw-Hill, 1968.

27. H. K. Gummel, IEEE Trans. on Elec. Dev., ED-11, 455 (1964).

28. G. D. Hachtel, R. C. Joy, and J. W. Colley, Proc. IEEE, 60, pp. 86-98, (Jan. 1972).

29. F. A. Lindholm, IEEE Trans. on Circuit Theory, CT-18, pp. 122-127 (Jan. 1971).

30. J. A. Narud and M. J. Callahan, IEEE Trans. on Circuit Theory, CT-11, 312 (1964).

31. F. A. Lindholm and D. J. Hamilton, Solid State Electronics, 7, (1964).

32. W. H. Howard, "Device Modelling for Computer-Aided Circuits Analysis," 1969 Wescon Technical Papers, S. F. August, 19-22, 1969, Session 23.

33. F. A. Lindholm, S. W. Director, and D. L. Bowler, IEEE J. of Solid State Cir., SC-6, p. 213-222, (Aug. 1971).

34. J. M. Early, Proc. I.R.E., 40, 1401 (1954).

35. T. W. Collins, Proc. IEEE, 57 840 (1969).

36. H. C. Josephs, Proc. IEEE, 55, 535 (1967).

37. H. K. Gummel and B. R. Chawla, IEEE Trans. on Elec. Devices, ED-18, 178 (1971).

38. L. Nagel and R. Rohrer, IEEE J. Solid State Circuits, SC-6, 166, (1971).

39. J. Logan, Proc. IEEE, 60, 335 (1972).

40. F. A. Lindholm and D. J. Hamilton, Proc IEEE, 59, 1377-1378, (Sept. 1971).

41. F. A. Lindholm, P. Rohr, and D. J. Hamilton, Proc IEEE, 60, 335 (1972).

42. F. A. Lindholm, IEEE Trans. Nuc. Science, NS-18, p. 206-211, (1971).

43. W. G. Howard, "Device Modelling for Computer-Aided Circuit Analysis," 1969 Wescon Technical Papers, S. F., August 19-22, 1969, Session 23.

44. H. K. Gummel, B.S.T.J., 49, 115 (1970).

45. L. E. Clark, IEEE Trans. Elec. Dev., ED-17, 661, (1970) (Contains extensive review on prior work.).

46. D. L. Bowler and F. A. Lindholm, IEEE Trans. on Electron Devices, ED-20, 257 (1973). (Contains extensive review of prior work.).

47. H. C. Poon and J. C. Meckwood, IEEE Trans. on Elec. Dev., 19, 90-97 (Jan. 1972).

48. J. R. Hauser, IEEE Trans. on Elect. Devices, ED-11, 238 (1964).

49. H. N. Ghosh, IEEE Trans. on Electr. Devices, ED-12, 513 (1965).

50. H. Groendijk, IEEE Trans. on Elec. Devices, 329 (March 1973).

51. A. van der Ziel and D. Agouridis, Proc IEEE, 54, 441 (1966).

52. J. Lange, IEEE Trans. on Elec. Div., Dec. 1971.

53. J.A. Archer, Solid State Elec., 15, 249-258 (1972).

54. N. J. Gri, "Microwave Transistors from Small Signal to High Power", Microwave Journal, Feb. 1971.

55. A. B. Macnee and R. J. Talsky, IEEE Journal of Solid State Circuits, Aug. 1972.

56. D. P. Kennedy and P. C. Murley, "Steady State Mathematical Theory for the Insulated Gate Field Effect Transistor, "IBM Journal, 17, no.1.

57. D. Vandorpe,J. Borel, G. Merckel, and P. Saintol, Solid State Elec., 15, 547-557 (1972).

58. M. Reiser, "Large-Scale Numerical Simulation in Semi-Conductor Device Modelling," Computer Methods in Applied Mechanics and Engineering, 1 (1972), p. 17-38.

59. M. Reiser and P. Wolfe, "Computer Study of Sub-micrometer F.E.T.S.," Electronic Letters, 8, no. 10.

60. F. A. Lindholm, IEEE Journal of Solid State Circuits, vol. SC. 6, Aug. 1971.

61. J. Meyer, "MOS Models and Circuit Simulation," RCA Review, 32 (Mar. 1971).

62. F. S. Jenkins, E. R. Lane, W. W. Lottin, and W. S. Richardson, IEEE Trans. Circuit Theory, CT-20, 649 (1973).

63. C. T. Sah, Phys. State Sol. (a) 7, 541 (1971).

64. C. T. Sah, Solid State Elec., 13, 1547 (1970).

65. C. T. Sah, Electronics Letters, 8, 88 (Feb. 24, 1972).

66. F. Hennig and C. T. Sah, Solid State Elec., 16, 1081-1083 (1973).

67. C. T. Smiley, L. D. Yao, and C. T. Sah, Solid State Elec., 16, 895 (1973).

68. J. P. Raymond, D. N. Pocock, and R. E. Johnson, Final Reprot, AFWL Contract F29601-70-C-0020, (1970).

69. J. P. Raymond, D. N. Pockock, and D. E. Meyerhoff, Final Report, contract USIVRDL N00228-65-C-0398, Report USNRDL-TRC-69-22, May 1969.

70. See, for example, past issues of the IEEE Transactions on Nuclear Sciences.

71. Charles Wilson and co-workers, Bell Telephone Laboratories, Murray Hill, New Jersey, U.S.A.

72. D. D. Kleppinger and F. A. Lindholm, Solid State Electronics, 14, 199 (1971).

73. D. D. Kleppinger and F. A. Lindholm, Solid State Electronics, 14, 407 (1971)

74. R. J. Van Overstraeten, H. J. DeMan and R. P. Mertens, IEEE Trans., ED-20, 290 (1973).

75. H. J. DeMan, R. P. Mertens and R. Van Overstraeten, Electronics Letters, Vol. 9, No. 8/9 (1973).

76. R. P. Mertens, H. J. DeMan, and R. J. Van Overstraeten, IEEE Trans. on Electron Devices, ED-20, 772, (1973).

77. M. S. Mock, Solid State Electronics, 16, 1251 (1973).

78. H. DeMan and R. Mertens, Digest of 1973 International Solid State Circuits Conference, 16, (1971).

79. P. Rohr, F. A. Lindholm, and K. R. Allen, "Questionability of Drift-Diffusion Transport in the Analysis of Small Semiconductor Devices," Solid State Electronics, to appear; see section immediately following in this report.

80. D. N. Pocock, M. G. Krebe, J. P. Raymond, and W. W. Chang, Final Report, AFWL TR-71-183, NCL 71-52R, Dec. 1971.

81. D. P. Kennedy, private communication.

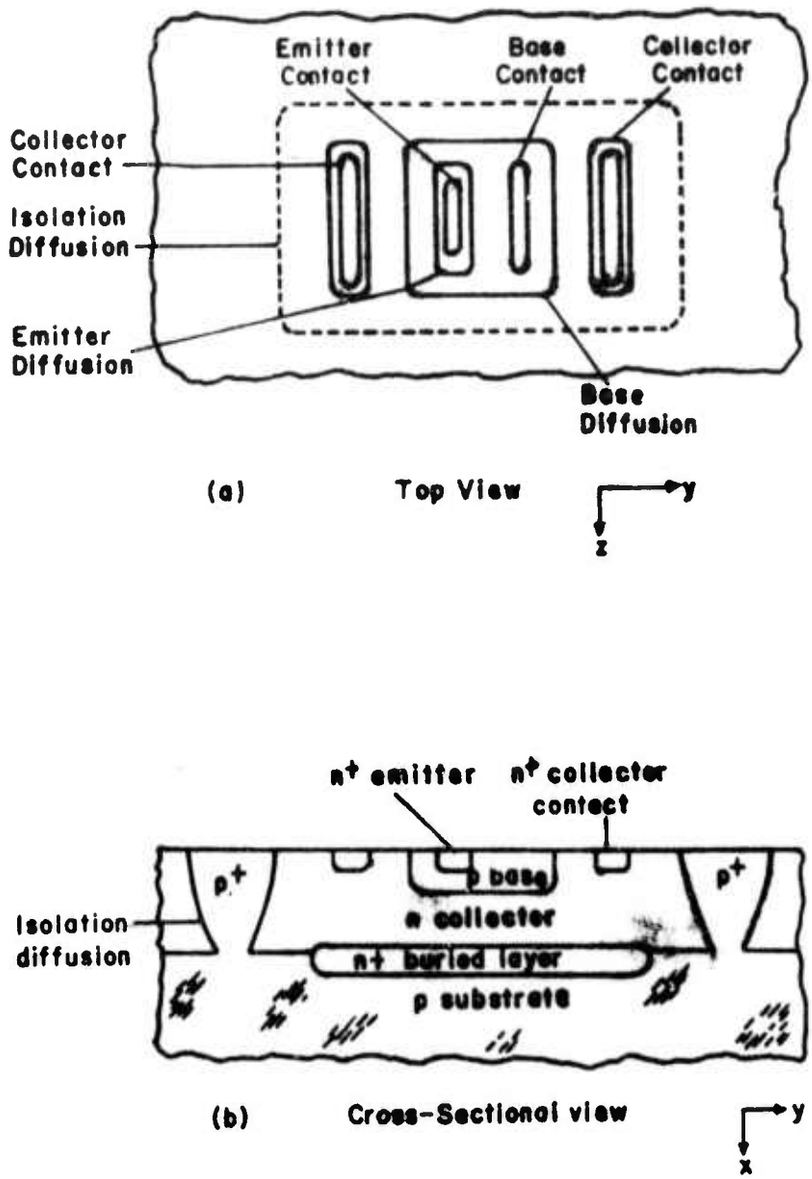


Fig. 1 Indicating the Complexity of the Boundary Value Problem Caused by the Structure of a Bipolar Transistor.

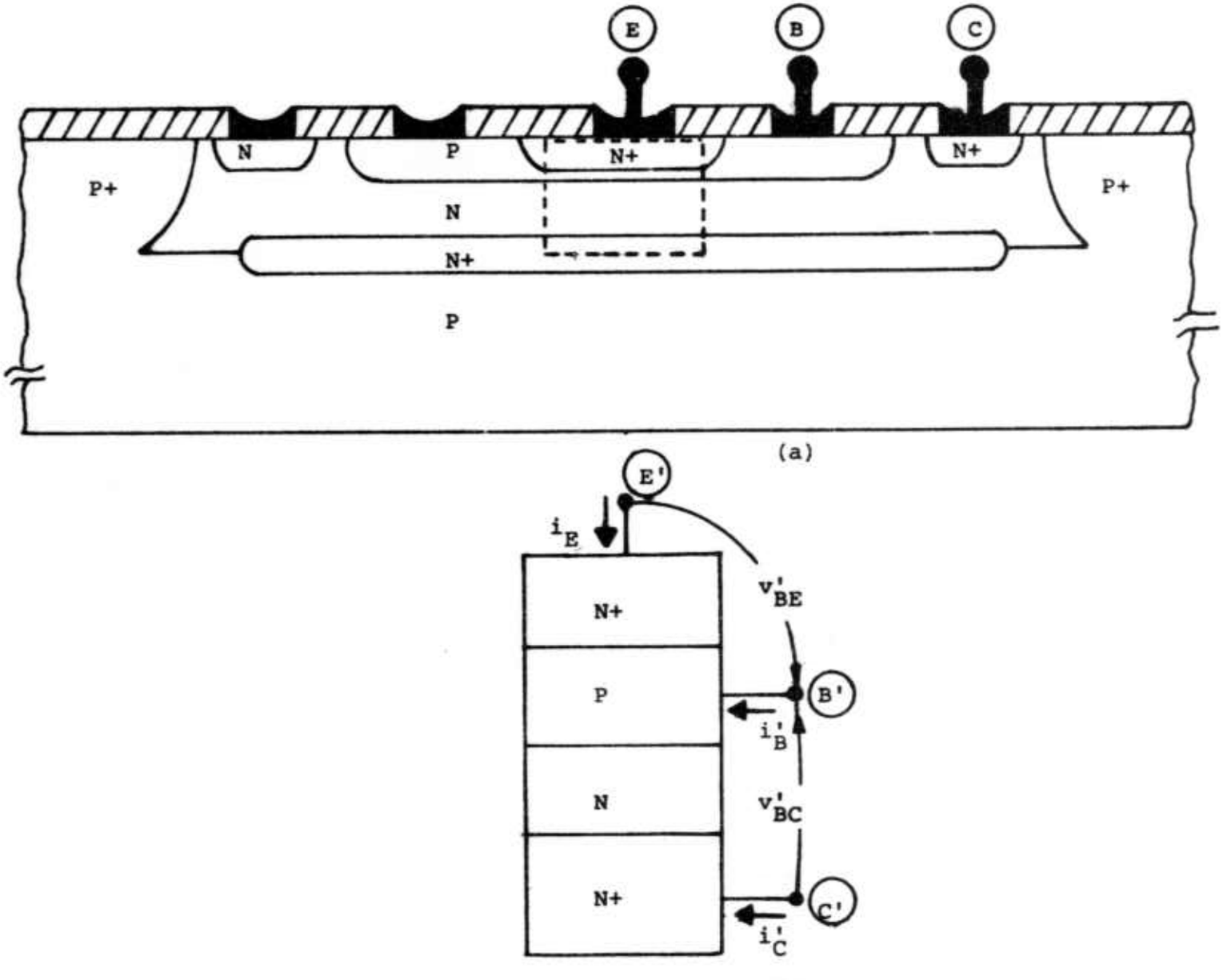


Fig. 2 Reducing dimensions by sectioning into intrinsic and extrinsic parts: (a) the sectioning, (b) the intrinsic part.

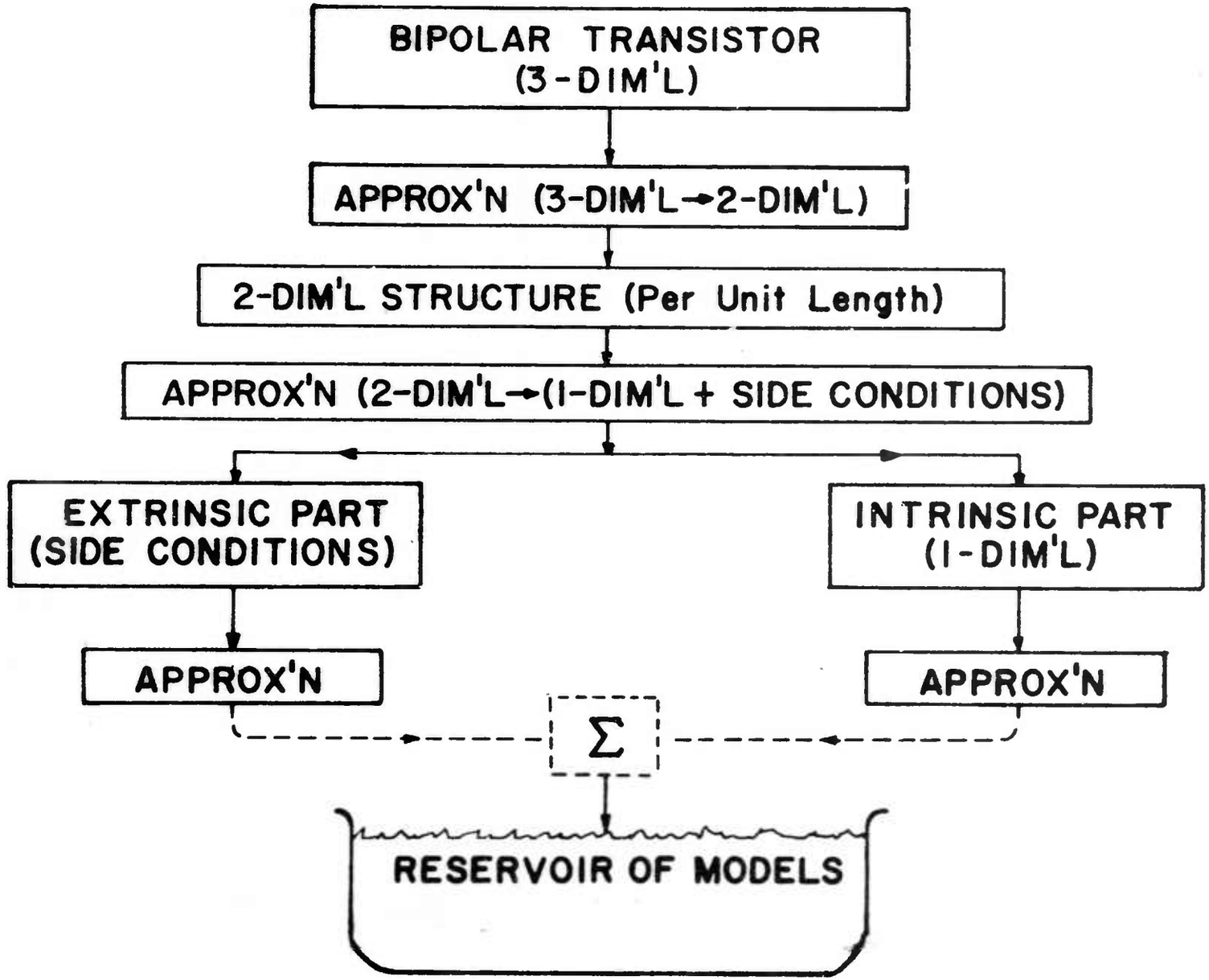
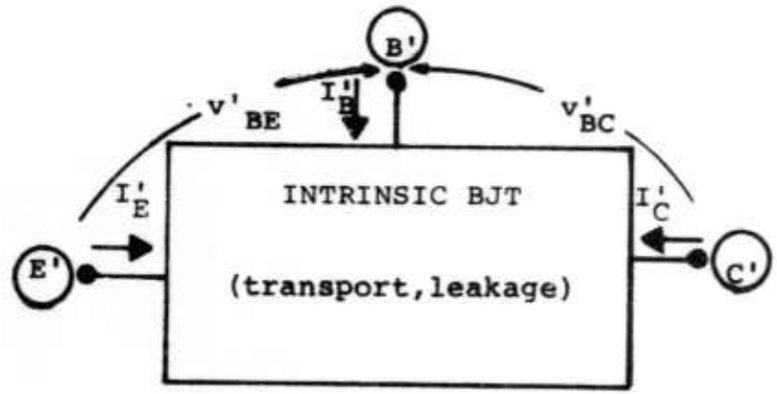
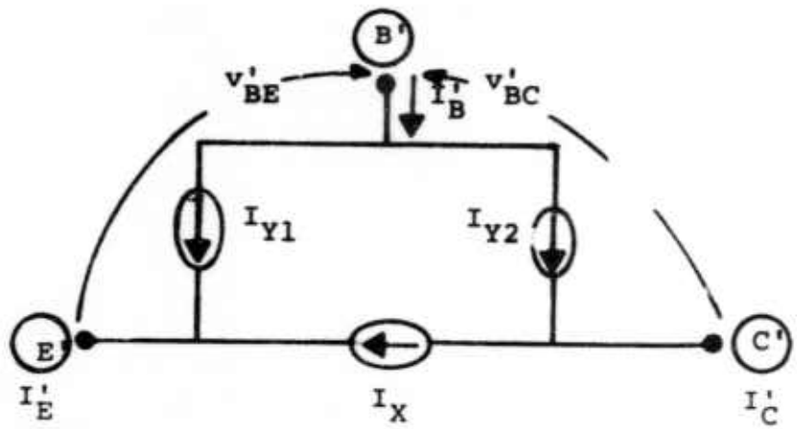


Fig. 3 Approximations. The upper half of the figure conveys the two reduction-of-dimension approximations. The lower half indicates additional approximations used to get reservoir of circuit models now available. These approximations will be discussed subsequently.

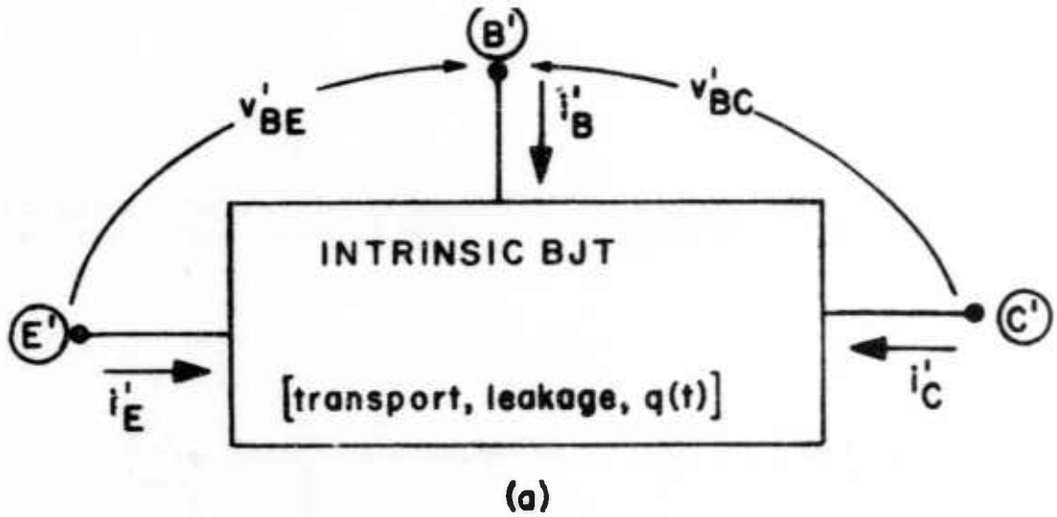


(a)

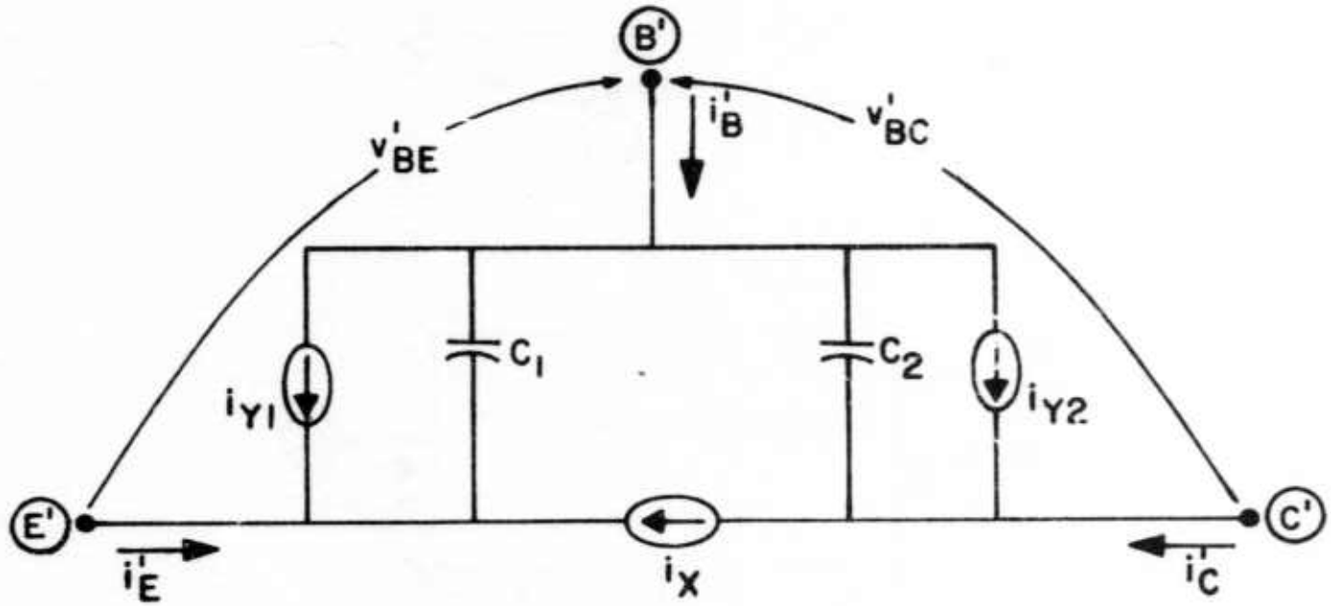


(b)

Fig. 4 General model for DC excitation: (a) indicating dominant processes within the intrinsic part; (b) corresponding equivalent circuit diagram.



(a)



(b)

Fig. 5 General model for time-varying excitation: (a) indicating dominant processes within the intrinsic part; (b) corresponding equivalent circuit diagram.

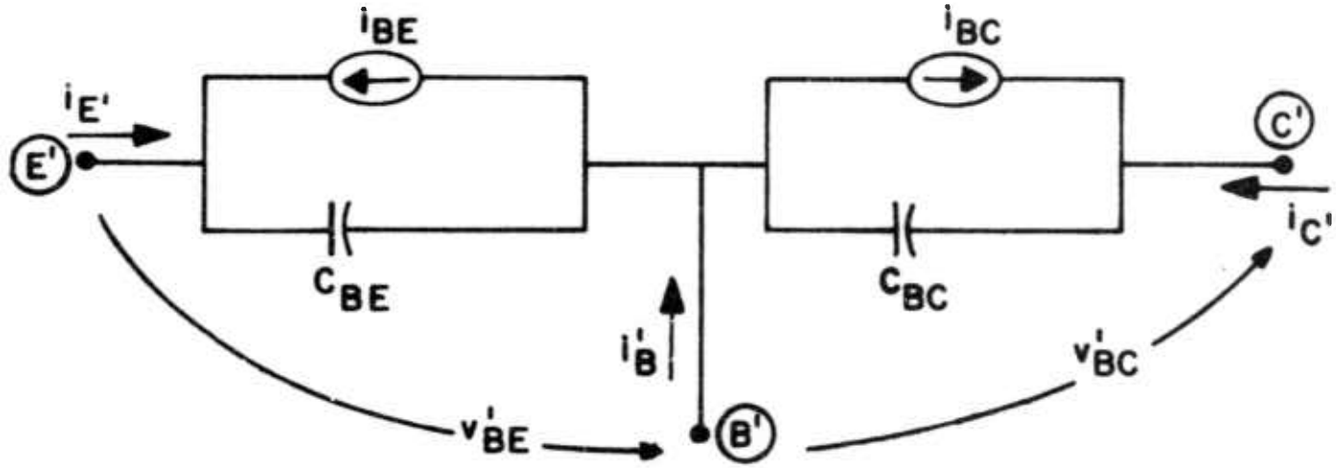


Fig. 6 The topology of either form of the Ebers-Moll model and of the class of models derived directly from it.

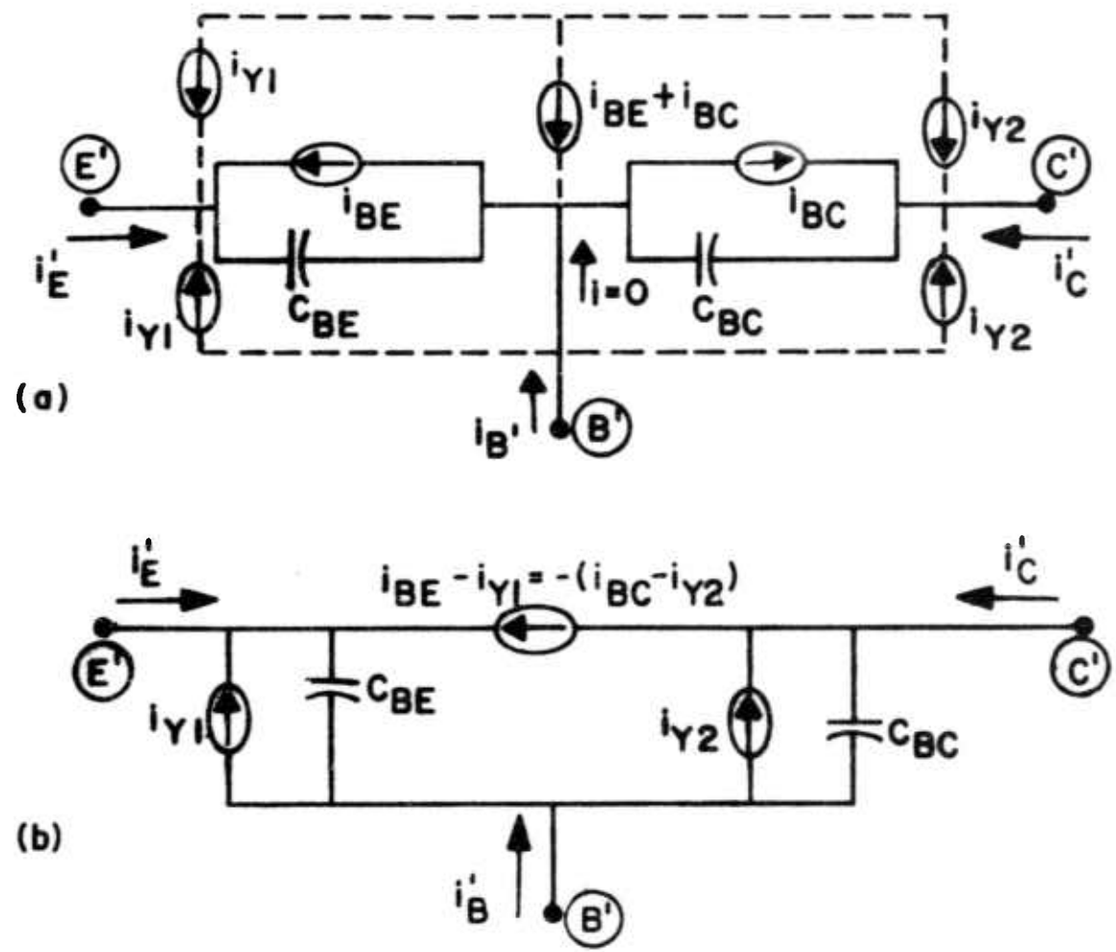
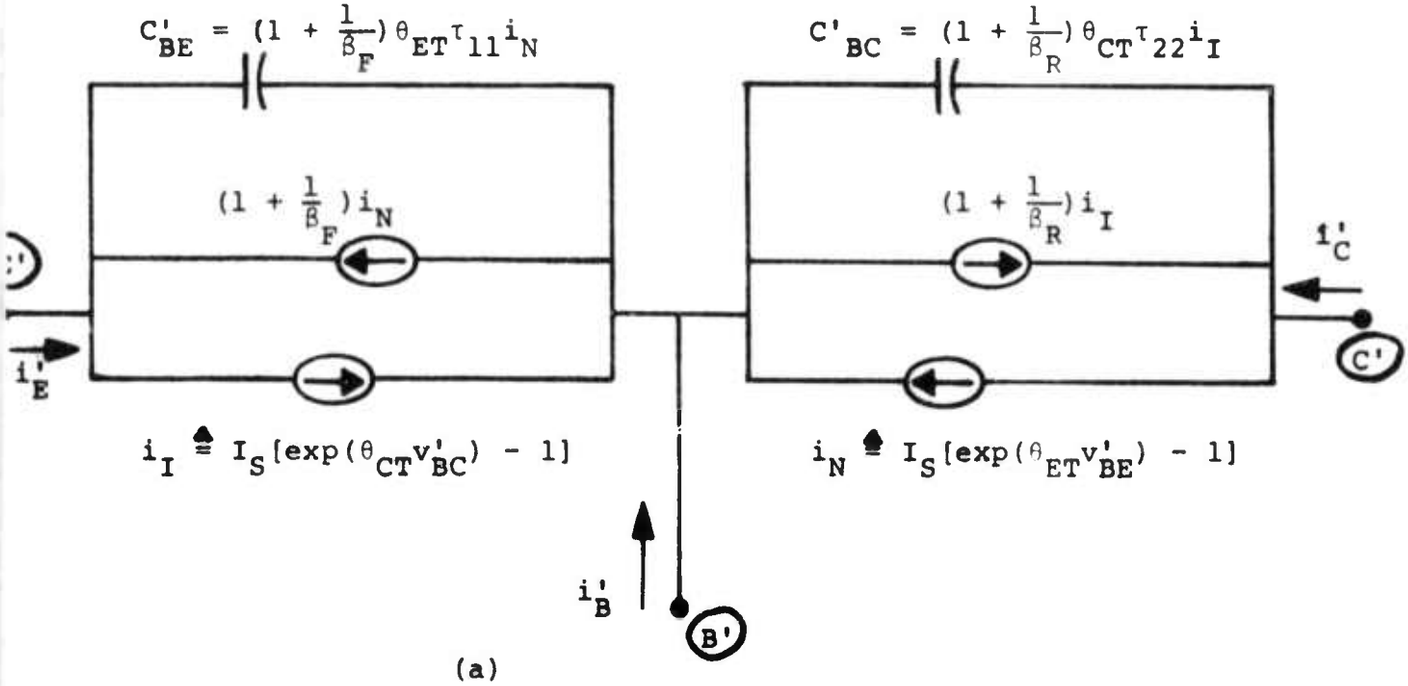
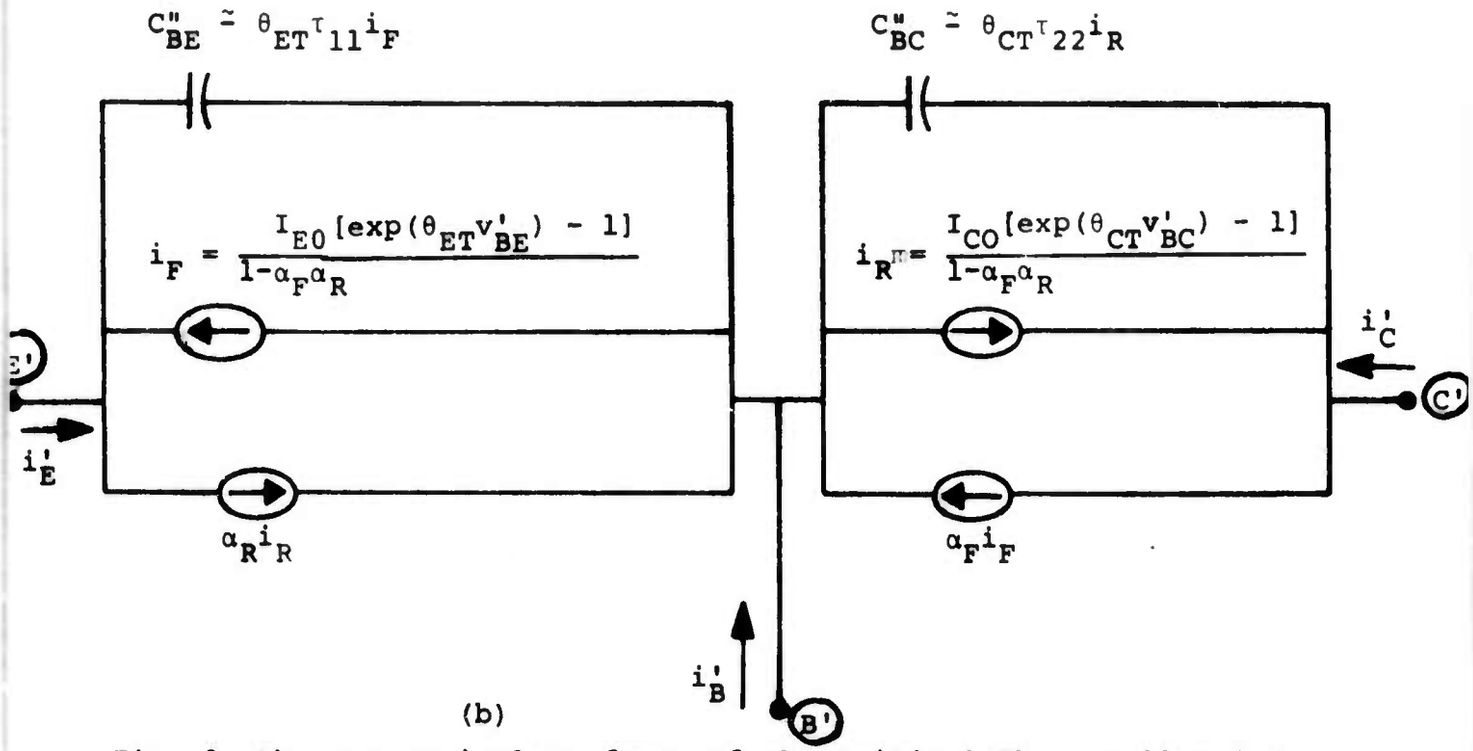


Fig. 7 Indicating an alternative way to demonstrate Eqs. (20) and (21): (a) starting with the circuit diagram of Figure 6 (solid lines), add current sources (dashed lines) that make the topology resemble that of Figure 5 while leaving the terminal behavior unchanged. The uppermost five current sources are then replaced by a single source with a value that again leaves the terminal behavior unchanged. (b) the result of these manipulations has the topology of the general circuit diagram of Figure 5 and values of current sources and capacitors consistent with Eqs. (20) and (21).



(a)



(b)

Fig. 8 The two equivalent forms of the original Ebers-Moll model: (a) the transport form; (b) the equality of these two circuit diagrams may be shown in a straightforward way be defining

$$I_S = \alpha_R I_{CO}/(1-\alpha_F\alpha_R)$$

and using the relation

$$\alpha_F I_{EO} = \alpha_R I_{CO}$$

which Ebers and Moll derived in their original paper.

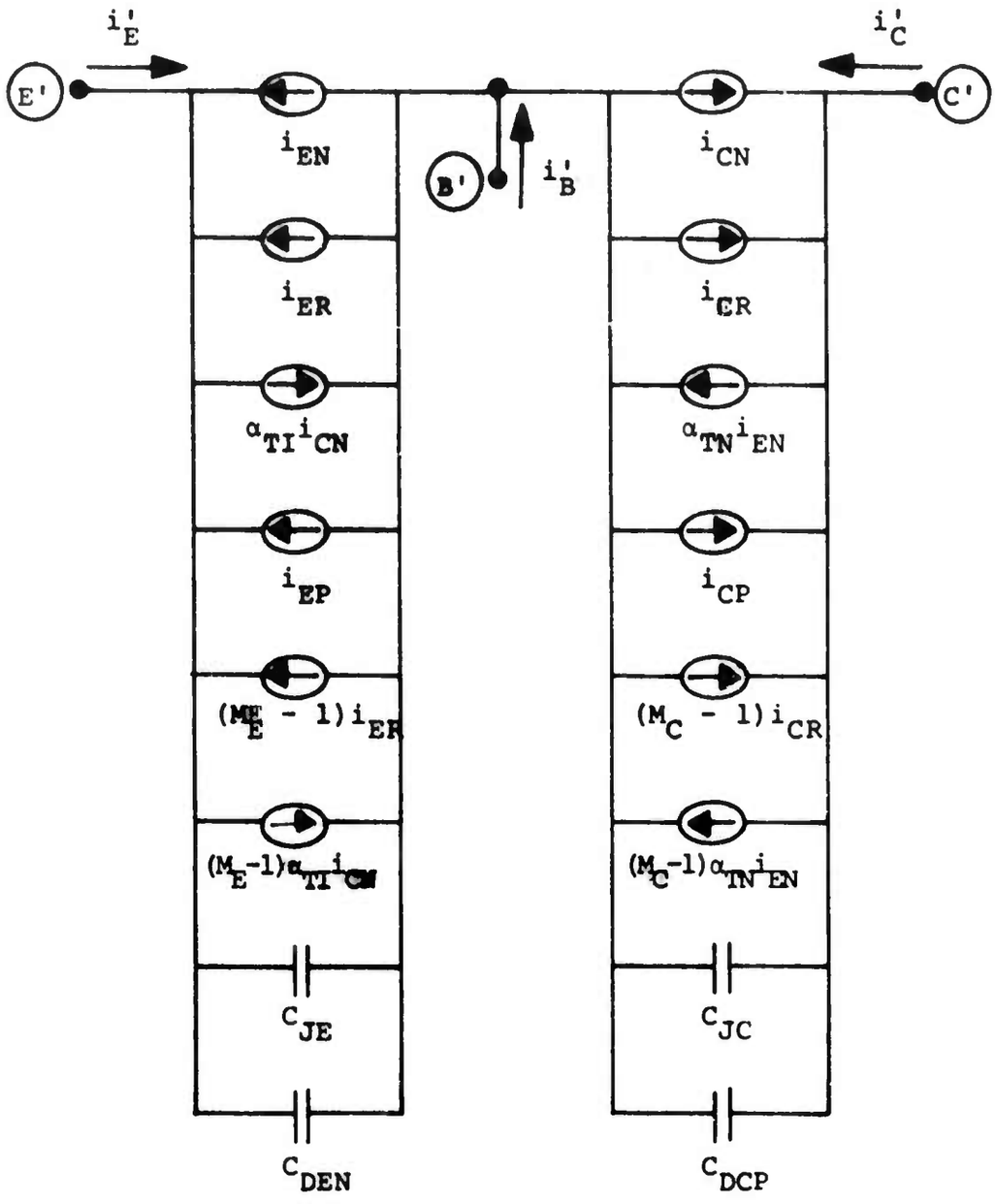


Fig. 9 Fossum's expandable model for the intrinsic bipolar transistor.

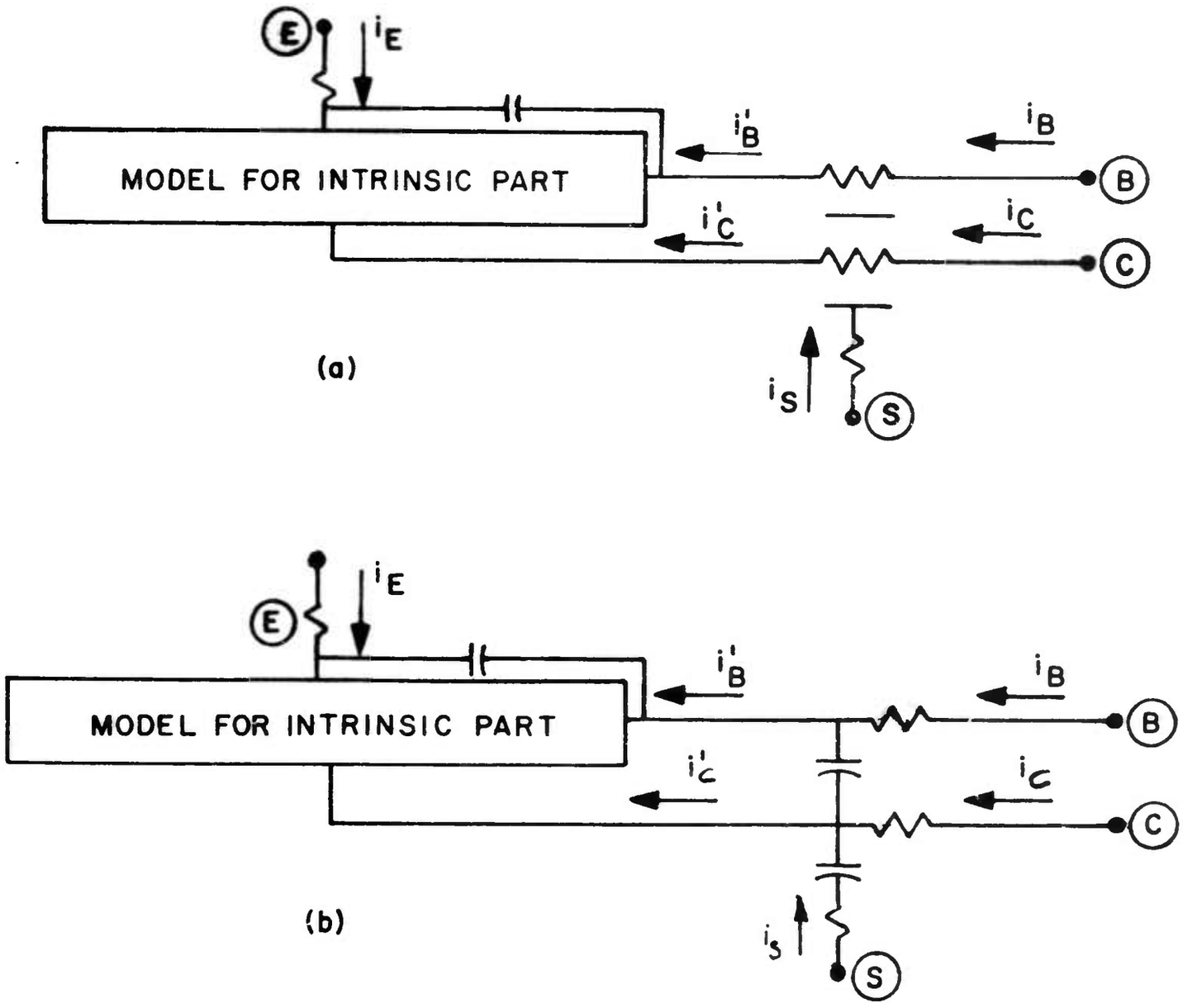


Fig. 10 Modeling the extrinsic part of the bipolar transistor (see Fig. 2).

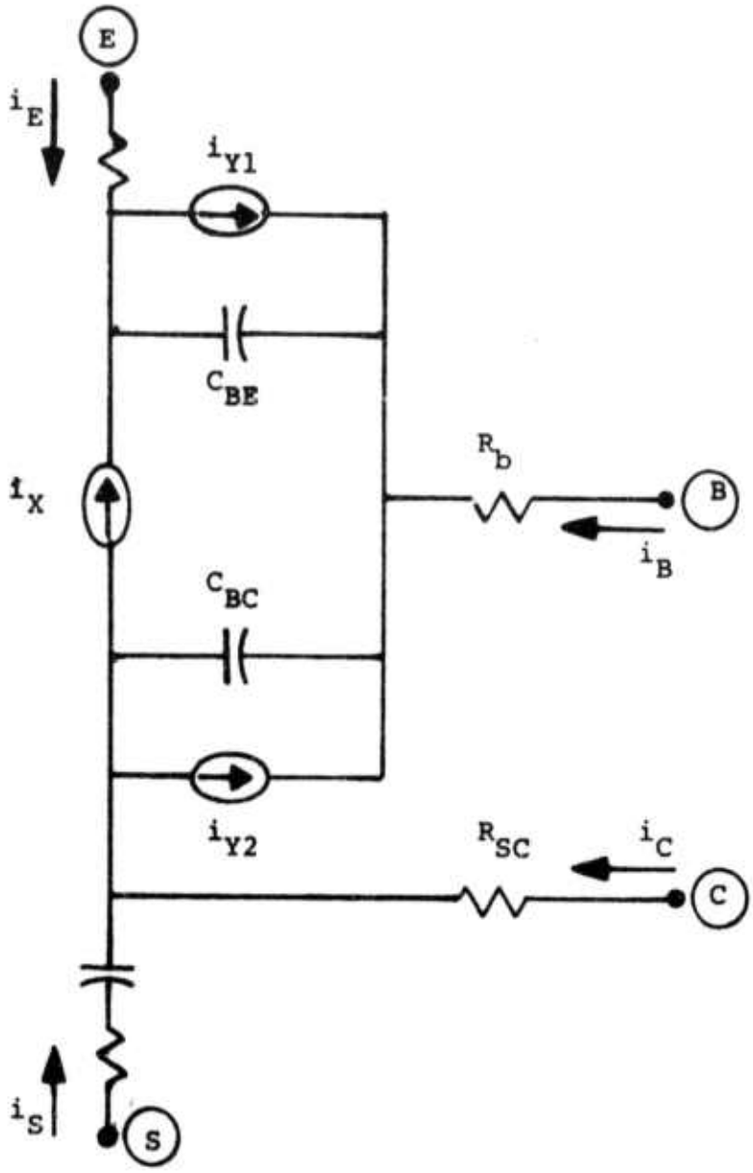


Fig. 10(c) Combination of the models of Figs. 5(b) and 10(b) yields a model for the entire bipolar transistor.

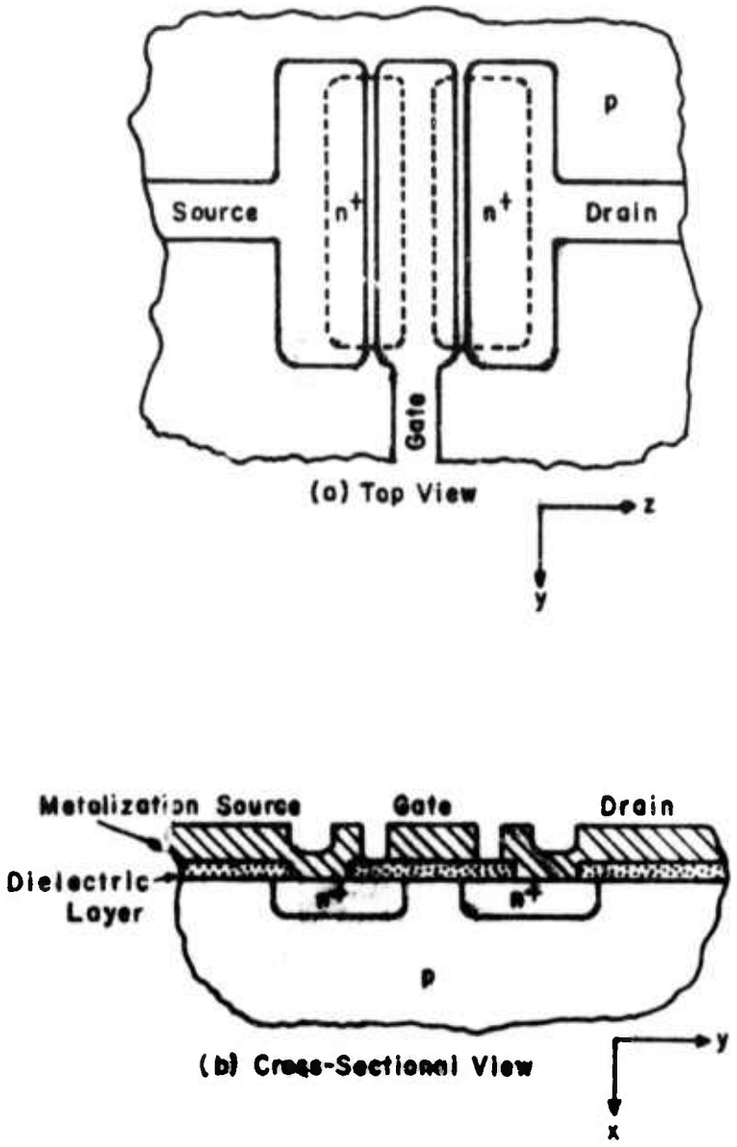


Fig. 11 IGFET or MOST; illustrating the pertinent three dimensional boundary value problem.

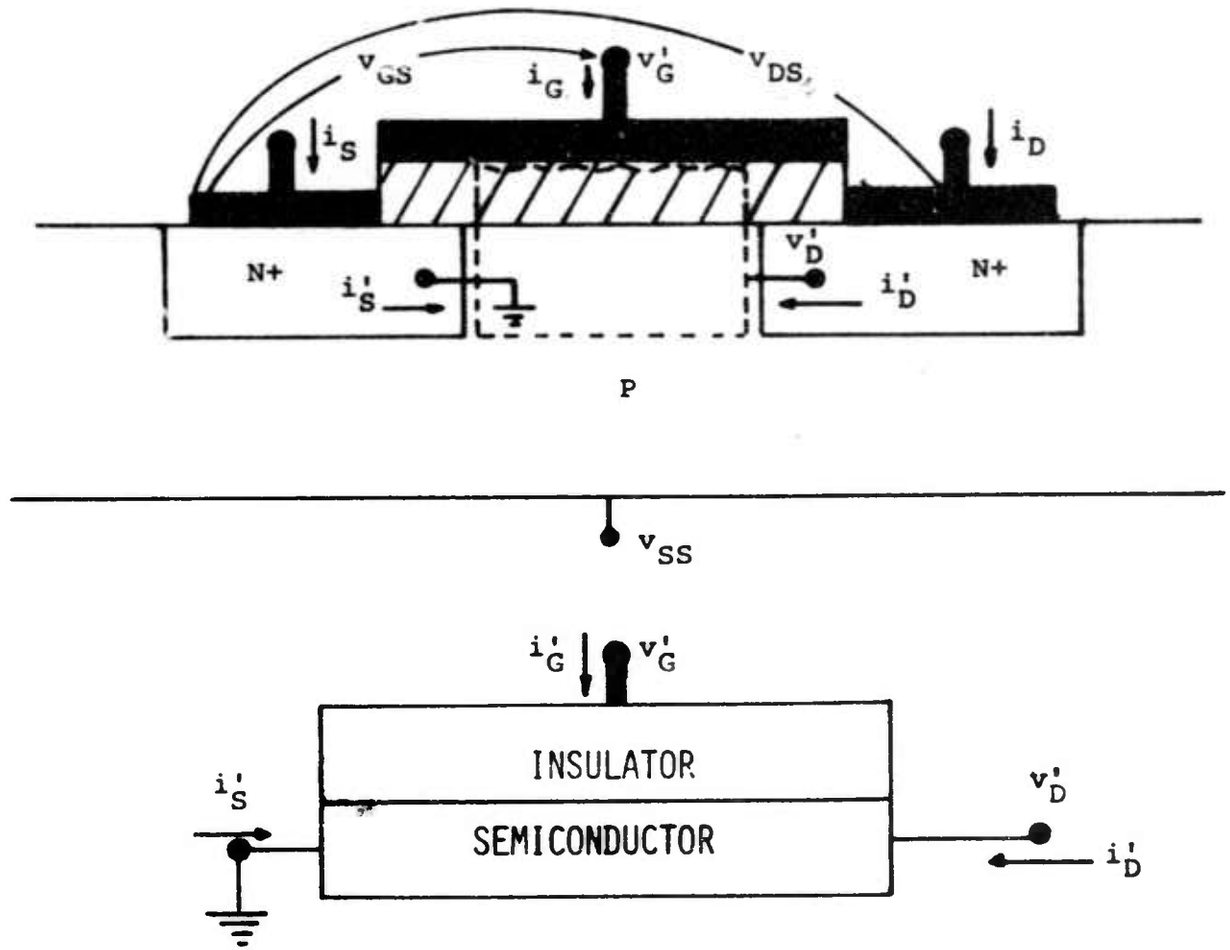


Fig. 12 Sectioning of IGFET into Extrinsic & Intrinsic Parts. The intrinsic part lies within the dotted lines.

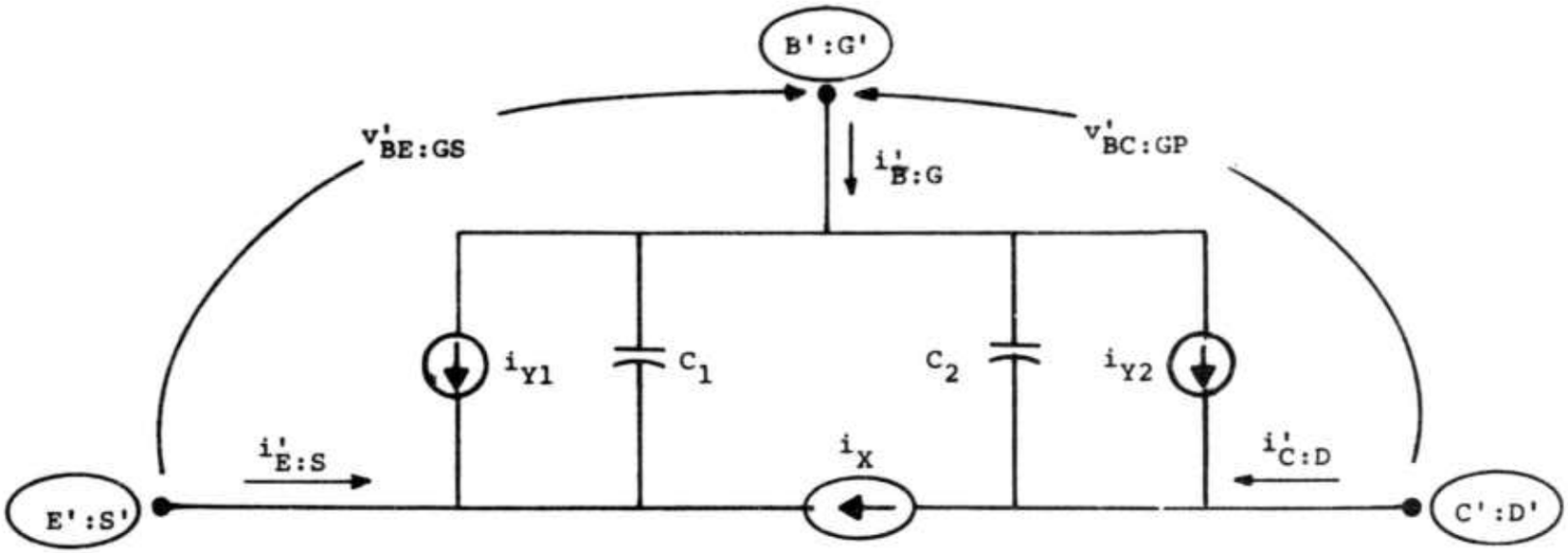


Fig. 13 Equivalent circuit diagram applying to large-signal dynamic model for intrinsic BJT or FET.

III. Questionability of Drift-Diffusion Transport in the Analysis of Small Semiconductor Devices (P. Rohr, F. A. Lindholm, and K. R. Allen)

INTRODUCTION

Mathematical studies pertinent to the analysis and design of such semiconductor devices as bipolar and field-effect transistors rely fundamentally upon equations that express the flow and conservation of carriers,

$$\vec{J}_n = q\mu_n n\vec{E} + qD_n \nabla n \tag{1}$$

$$\vec{J}_p = q\mu_p p\vec{E} - qD_p \nabla p \tag{2}$$

$$\frac{\partial n}{\partial t} = -U + \frac{1}{q} \nabla \cdot \vec{J}_n \tag{3}$$

$$\frac{\partial p}{\partial t} = -U - \frac{1}{q} \nabla \cdot \vec{J}_p \tag{4}$$

$$\vec{J} = \vec{J}_p + \vec{J}_n \ , \tag{5}$$

combined with the Maxwell equations of electromagnetism. These equations were used by Shockley [1] in 1949 in the first theoretical description of pn junction behavior and were unified in 1950 by van Roosbroeck [2]. They have served as a basis for traditional semiconductor device theory which, through employment of approximations, describes device operation by analytical expressions. These equations now serve as a basis for the computer-aided design and analysis of semiconductor devices and integrated circuits.

Reasons of economy, reliability, and performance often dictate that the number of devices contained in a single integrated circuit be as large as possible, which requires that devices be designed as small as possible. Because of this general trend in the semiconductor technology toward ever smaller devices, questions about Eqs. (1)-(5) now arise.

Here we shall limit consideration to Eqs. (1) and (2), suggesting that the small dimensions even of present-day devices may stretch the

theory underlying these equations beyond the bounds of applicability. We shall focus attention mostly on the diffusion component, though in a broad sense our remarks will apply also to the drift component.

QUALITATIVE GROUNDS FOR QUESTIONABILITY

The expression describing flow by diffusion contained in Eqs. (1) and (2) can be derived by starting from several different models for particle behavior: among these are the model of random flights treated by Markoff's method [3], and the relaxation-time model used in the Boltzmann transport equation [4]. As a common thread among the derivations, each model assumes, either tacitly or explicitly, a volume of solid large enough to contain many scattering centers. Under the influence of external excitation, the many collisions thus assumed in these models tend to cause the carriers to travel tortuous paths between any two points under consideration in the solid and tend to maintain carrier behavior similar to that present in thermal equilibrium.

In the derivation from the Boltzmann transport equation, one commonly employs a relaxation time approximation [4]. Thus one assumes that the numerous collisions occurring within the solid will tend to return the carrier distribution function to its equilibrium form at a rate due to collisions that varies in proportion to its deviation from the equilibrium form. When critical regions of a semiconductor device shrink to dimensions of the same order as a mean free path between collisions, the relaxation-time approximation becomes questionable because of insufficient scattering of carriers.

Conventional fabrication techniques produce high-performance bipolar transistors with metallurgical base widths less than 1000 Å [5]. For a transistor with a metallurgical basewidth of 1000 Å, conventional pn junction theory, which includes Eqs. (1)-(5), predicts an electrical (quasi-neutral) base width of approximately 300 Å. This is of the same order as the mean free path between collisions in the base. Thus, of the order of one collision may occur as the average carrier crosses the electrical base region.

The main purpose in what follows here is to indicate some consequences of continuing to describe carrier transport by diffusion in semiconductor devices containing critical regions so small that the validity of this description becomes questionable. To do this, we define a hypothetical transistor whose properties apply in the limiting case of small dimensions.

THE HYPOTHETICAL COLLISIONLESS TRANSISTOR

In accord with traditional semiconductor device theory, we define the collisionless transistor to comprise quasi-neutral regions, substantially empty of space charge, separated by junction barrier regions, filled with space charge. This sectioning, together with the corresponding energy-band configuration, is shown in Fig. 1. The electrical (quasi-neutral) base region has width W, and its edge nearest the emitter is located at $x = 0$.

In contrast to traditional semiconductor device theory, we define the collisionless transistor to have the following related properties:

(a) the electrical base region is so thin that carriers crossing it experience no collisions; and

(b) outside the electrical base, in the emitter and collector, and in the junction barrier regions, the carriers experience numerous collisions.

Thus we section not only into quasi-neutral and space-charge regions but also into collisionless regions and regions in which many collisions occur. By definition, sharp boundaries separate adjoining regions.

In the limiting sense, this further sectioning recognizes trends that occur in small transistors. As was noted earlier, transistors now exist that have electrical base widths of the same order as the mean free path between collisions, and a carrier crossing the electrical base will indeed experience little scattering. In the collector and emitter, however, much scattering can be expected because, on the one hand, the collector is the thickest region in the intrinsic transistor, and, on the other hand, the high doping concentration in the emitter tends to depress both the mobility [6] and the mean free path more than an order of magnitude below their values in the electrical base.

We limit attention solely to the current density J due to electrons flowing in the x-direction. First, consider the component J_+ in the dc steady state arising from electrons that surmount the emitter barrier and cross the electrical base from left to right. In the dc steady state, J_+ is independent of x and is calculated most easily at x = 0:

$$J_+ = qn_+(0)v_+(0) \tag{6}$$

Here $n_+(0)$ denotes the number density of electrons having enough energy to surmount the emitter barrier, and $v_+(0)$ denotes the mean velocity of electrons crossing from left to right.

If the applied voltage is small enough that the quasi-Fermi levels remain practically constant across the emitter barrier region [7], then the electrons in the conduction band in this region are described by a distribution function f that is nearly spatially independent. This implies that these electrons are practically in equilibrium among themselves. Expressions for $n_+(0)$ and $v_+(0)$ then derive from straightforward integrations of the distribution function f. We assume f to be the Boltzmann tail of a Fermi-Dirac distribution. For the number density, the result of integration is

$$n_+(0) = n_n \exp[-q(\phi_{EB}-V_{BE})/kT] \tilde{=} n_{po}\exp(qV_{BE}/kT) \tag{7}$$

For the one-sided mean velocity $v_+(0)$ the result is [8, 9]

$$v_+(0) = \left(\frac{m^*}{2\pi kT}\right)^{1/2} \int_0^\infty v_x e^{-\frac{m^*v_x^2}{2kT}} dv_x = \left(\frac{kT}{2\pi m^*}\right)^{1/2} \tag{8}$$

Using analogous reasoning and notation, we calculate the component,

$$J_- = qn_-(W)v_-(W) \ , \tag{9}$$

arising from electrons that surmount the collector barrier and cross the electrical base from right to left. We find

$$J = J_+ - J_- = [J_S]_{direct}[\exp(qV_{BE}/kT) - \exp(qV_{BC}/kT] \tag{10}$$

for the total current density of electrons. In Eq. (10),

$$[J_S]_{direct} = qn_{po}[kT/2\pi m^*]^{1/2} \quad (11$$

denotes the saturation current density applying for direct-transition (no-collision) transport across the electrical base.

PREDICTIONS OF TRADITIONAL DEVICE THEORY

Conventional theory assumes that throughout the transistor current flows as a result of drift and diffusion, in accord with Eqs. (1) and (2). Under low injection conditions [10], consistent with the near constancy of the quasi-Fermi levels in the barrier regions required earlier, minority carriers in the electrical base move predominantly by diffusion,

$$J(x) = qD \frac{dn}{dx} , \quad (12)$$

provided, as Fig. 1 implies, negligible electric field is built into the base. For lifetimes long compared to the transit time for the electrical base, the electron current density becomes [10]

$$J = [J_S]_{diffusion}[\exp(qV_{BE}/kT) - \exp(qV_{BC}/kT)] \quad (13)$$

$$= (qD/W)\ [n(0) - n(W)].$$

In Eq. (13),

$$[J_S]_{diffusion} = qDn_{po}/W, \quad (14)$$

denotes the saturation current density applying for diffusion transport across the electrical base; and, in contrast with $n_+(0)$ and $n_-(W)$, the symbols n(0) and n(W) denote <u>total</u> electron densities.

COMPARISON

The distinction between Eq. (10), which applies for the limiting case of small base widths, and Eq. (13), which derives from conventional transistor theory, lies in the functional dependence of the saturation current densities, defined in Eqs. (11) and (14). As one major difference, Eq. (14) predicts dependence on the electrical base width W while Eq. (11) predicts absence of such dependence. Fig. 2(a) illustrates this by plotting Eq. (11), for direct transitions, as the asymptote for small W and Eq. (14), for diffusion, as the asymptote for large W. The dashed

curve in Fig. 2(a) proposes qualitatively the non-asymptotic dependence of saturation current on electrical base width. In Fig. 2(b), we indicate a method for experimental determination of the saturation current, $I_S = J_S \times$ Area, to suggest how these theoretical dependences might be put to experimental test.

A second related distinction lies in the different dependences on temperature predicted for the current density. This is revealed by the ratio of saturation current densities,

$$[J_S]_{direct} / [J_S]_{diffusion} = [kT/2\pi m^*]^{1/2} / [\mu kT/qW]. \quad (15)$$

The ratio varies in proportion to T if lattice scattering predominates in determining the theoretical temperature dependence of the mobility μ in the electrical base region, and in inverse proportion to T^2 if impurity-ion scattering predominates [11].

A third difference appears in the distribution of the mobile carriers across the electrical base. In the collisionless transistor, the distribution is constant, independent of the spatial coordinate x:

$$n(x) = n_+(0) + n_-(W)$$

$$= n_{po} \left[\exp(qV_{BE}/kT) + \exp(qV_{BC}/kT)\right]. \quad (16)$$

This occurs because electrons emitted from one junction move toward the other at a constant velocity, there being no forces in the electrical base of the collisionless transistor to change the velocity. In conventional transistor theory, however, the distribution is a line of constant slope:

$$n(x) = n(0) \left[1-x/W\right] + n(W) \left[x/W\right]$$

$$= n_{po}\left[(1-x/W)\exp(qV_{BE}/kT) + (x/W)\exp(qV_{BC}/kT)\right]. \quad (17)$$

The difference in these distributions implies certain consequences about transistor speed, which follow from the charge-control relation [12],

$$J = Q/\tau. \quad (18)$$

Here Q denotes the charge per unit area in transit and τ denotes the pertinent charge-control time constant (transit time). In forward-active operation, for example, operation at a given current density implies

$$[\tau]_{direct}/[\tau]_{diffusion} = [Q]_{diffusion}/[Q]_{direct}$$

$$= 2\,\frac{[J_S]_{diffusion}}{[J_S]_{direct}} \qquad (19)$$

$$= (2D/W)/[kT/2\pi m^*]^{1/2}$$

which is of the order of $10^{-5}/W$, with W measured in centimeters.

Fig. 3 displays the dependence of the transit time on the electrical base width W. For the limiting case of small W, in which flow by direct transistion dominates, the transit time varies in proportion to W. For the limiting case of large W, in which flow by diffusion dominates, the transit time varies in proportion to W^2. Intermediate values of W will produce a compromise in dependence, as the qualitative projection of the dashed curve of Fig. 3 indicates.

Similarly, the absence of collisions in the transistor base will cause the contribution of the base to the saturation storage time to depart from the predictions of conventional theory.

DISCUSSION

The foregoing comparison shows the difference in transistor behavior predicted by two models: one representing the limiting case of large dimensions, and the other the limiting case of small dimensions. The contrast in predictions, underscored in Figs. 2 and 3, suggests that analysis based on assumed diffusion could misguide the design of small semiconductor devices.

The two models lend themselves to straightforward calculations. Because of the idealizations made, however, neither model provides a wholly realistic description of transistor behavior, even for the two limiting cases they are proposed to represent. Each model ignores certain contributors to transistor behavior. To insure perspective, therefore, some further comment is necessary.

Both models make certain of the same idealizations:

(a) neglect of net recombination and generation in the junction barrier regions and at the surface [10];

(b) neglect of high injection [13];

(c) neglect of multi-dimensional flow, caused, for example, by such phenomena as emitter crowding [14] and high current modes in the collector [15];

(d) neglect of the warping of the quantum density of states in the emitter and the base that accompanies high doping [16]; and

(e) neglect of base-width modulation (Early effect) [17].

These idealizations are made for simplicity of discussion, to avoid obscuring the central intent of this paper.

For the model termed the collisionless transistor, we have made additional idealizations concerning the location of scattering events. We have assumed that outside the electrical base--in the emitter, collector, and junction-barrier regions--numerous collisions occur. In conflict with this assumption, note that the width of the emitter junction region of certain transistors in forward bias might shrink to dimensions of the same order as the mean free path. We have assumed that sharp boundaries separate collisionless regions from adjoining regions in which collisions occur. In fact, any such boundary would not be abrupt, but rather would probably better be described by a blur of thickness comparable to a mean free path. While these assumptions may introduce only small error for certain transistors, we have made them here again mainly to ease discussion; this is consistent with the intent of contrasting as simply as possible behavior deriving from diffusion transport with that from transport by direct transition.

Finally, the definition of the collisionless transistor contains the assumption that no collisions occur in the electrical base region, it being so thin. Calculations based on this assumption describe one limiting case--that in which the electrical base is thin relative to a mean free path--just as calculations based on the assumption of drift-diffusion transport correspond to the opposite limiting case--in which the electrical base is thick relative to a mean free path. For small transistors of the present day, neither of these asymptotes may provide an adequate description. The dashed curves in Figs. 2(a) and 3 project qualitatively the compromise in behavior occuring when the thickness of the electrical base becomes comparable to a mean free path. A quantitative treatment is intended for future papers, as are the consequences of removing the other assumptions named above.

Other studies have pointed to the need for reconsidering the applicability of drift-diffusion transport in the analysis of semi-

conductor devices, but from a viewpoint different from the one taken in the present paper. Persky [9] has noted that analysis based on Eqs. (1)-(5) can predict gradients of carrier concentrations that imply effective diffusion velocities exceeding the thermal velocity given in Eq. (8). In attempting to remedy this difficulty, he assumes an empirical modification of the current transport equations. Like the conventional expressions of Eqs. (1) and (2), the modification assumes that flow depends upon the gradient of the carrier concentration. Thus it fails to give the correct dependence in the limiting case of zero collisions, because, as pointed out in the discussion of Eq. (16) above, a concentration gradient cannot be sustained in a field-free collisionless region.

In this paper, our main attention has fixed upon the electrical base region of the bipolar transistor. Questions about the appropriateness of assuming flow by drift and diffusion, however, arise also for semiconductor devices other than the bipolar transistor: as an example, for such field-effect structures as MOS and Schottky-barrier (MES) transistors. Indeed, questions arise whenever a region critical to the behavior of a device has dimensions comparable to a mean free path between collisions. In future papers, we plan to discuss these broader implications.

LIST OF SYMBOLS

$\vec{J}_p$, $\vec{J}_n$	hole, electron current density
J_+, J_-	electron current density (+ for flow from left to right, - for flow from right to left)
J_S	saturation current density
$\vec{E}$	electric field
D_p, D_n	hole, electron diffusion coefficient
μ_p, μ_n	hole, electron mobility
q	electron charge
Q	total base charge in transit per unit area
p, n	hole, electron concentration
$n_+(0)$, $n_-(W)$	number densities of electrons surmounting barriers
n_n	electron density in the emitter at the edge of the barrier region
n_{po}	thermal-equilibrium electron density in the base region
U	net rate of recombination
t	time
τ	charge-control time constant (transit time) for forward active operation
$v_+(0)$, $v_-(W)$	one-sided mean carrier velocities
ϕ_{EB}, ϕ_{CB}	equilibrium barrier potentials
V_{BE}, V_{BC}	applied voltages across barrier regions
k	Boltzmann's constant
T	absolute temperature
m*	effective electron mass
W	electrical base width
E_F	Fermi level
E_C, E_V	conduction band, valence band edge

REFERENCES

1. W. Shockley, Bell Syst. Tech. J., 28, 435 (1949).

2. W. van Roosbroeck, Bell Syst. Tech. J., 29, 560 (1950).

3. S. Chandrasekhar, Rev. Mod. Phys., vol. 15, 1 (1943). This article is also reprinted in M. Wax, Selected Papers on Noise and Stochastic Processes, Dover Publications, New York (1954).

4. A. C. Smith, J. F. Janak, and R. B. Adler, Electronic Conduction in Solids, 315, McGraw-Hill, New York (1967).

5. (a) M. K. Barnoski and D. D. Loper, Solid State Electronics, 16, 433 (1973).

   (b) M. K. Barnoski and D. D. Loper, Solid State Electronics, 16, 441 (1973).

6. S. M. Sze, Physics of Semiconductor Devices, Wiley, New York, 40 (1969).

7. (a) C. T. Sah, IEEE Trans., ED-13, 839 (1966).

   (b) W. Shockley, Electrons and Holes in Semiconductors, 308, D. Van Nostrand, Princeton (1950).

8. F. Reif, Fundamentals of Statistical and Thermal Physics, 266, McGraw-Hill, New York (1965).

9. G. Persky, Solid State Electronics, 15, 1345 (1972).

10. A. S. Grove, Physics and Technology of Semiconductor Devices, John Wiley, New York (1967).

11. Ref. 10, 109.

12. (a) R. Beaufoy and J. J. Sparkes, ATE Journal, vol. B, 310 (1957).

    (b) J. J. Sparkes and R. Beaufoy, IRE Proc. 45, 1740 (1957).

    (c) E. O. Johnson and A. Rose, IRE Proc. 47, 407 (1959).

    (d) R. D. Middlebrook, IEE Proc. 106B, Suppl. 17, 887 (1959).

13. L. E. Clark, IEEE Trans., ED-17, 661 (1970).

14. (a) J. R. Hauser, IEEE Trans., ED-11, 238 (1964).

    (b) H. N. Ghosh, IEEE Trans., ED-12, 513 (1965).

15. D. L. Bowler and F. A. Lindholm, IEEE Trans., ED-20, 257 (1973).

16. (a) D. D. Kleppinger and F. A. Lindholm, Solid State Electronics, 14, 199 (1971).

(b) D. D. Kleppinger and F. A. Lindholm, Solid State Electronics, 14, 407 (1971).

(c) R. J. Van Overstraeten, H. J. DeMan and R. P. Mertens, IEEE Trans. ED-20, 290 (1973).

(d) H. J. DeMan, R. P. Mertens and R. Van Overstraeten, Electronics Letters, Vol. 9, No. 8/9 (1973).

17. J. M. Early, Proc. IRE, Vol. 46, 1141 (1952).

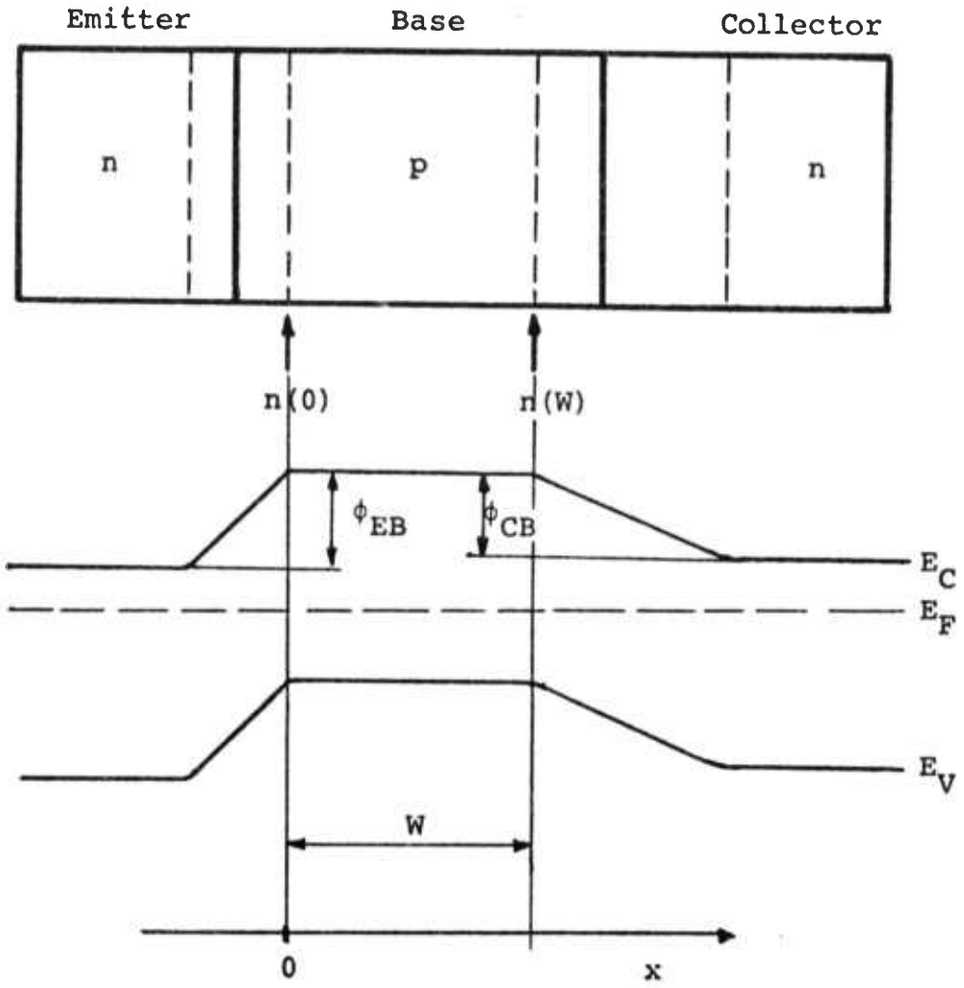


Figure 1 Illustrating the hypothetical, collisionless transistor. The energy-band diagram describes equilibrium conditions.

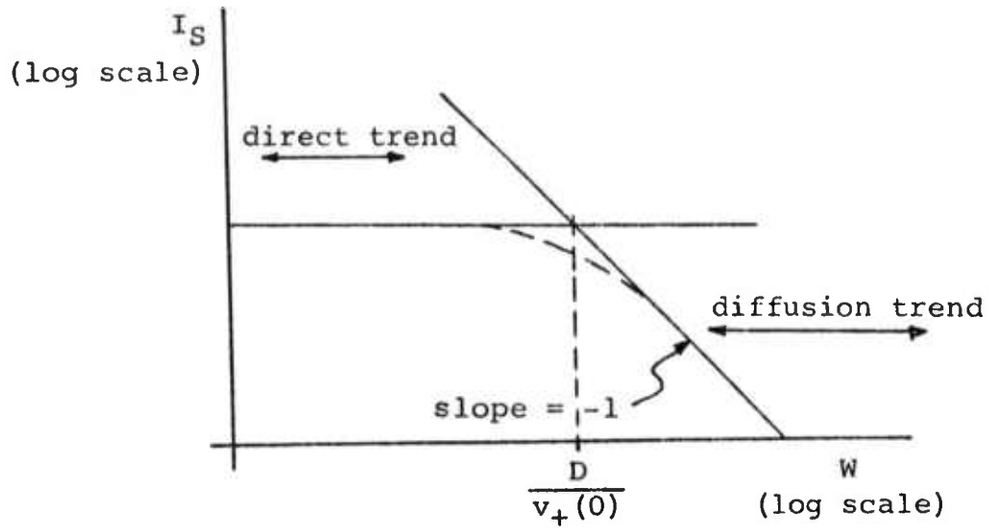


Figure 2(a) Saturation current I_S.

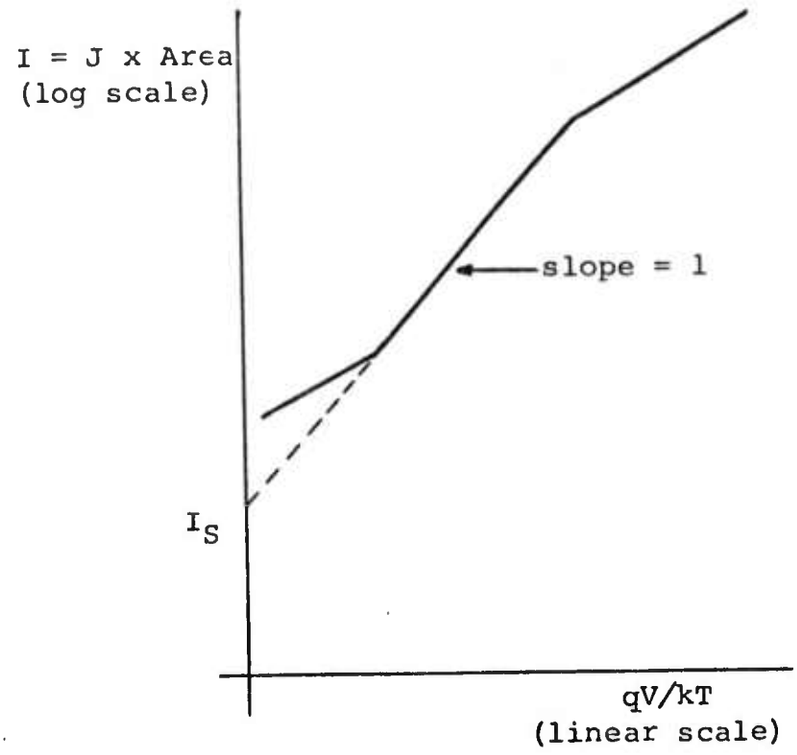


Figure 2(b) The saturation current I_S is found from extrapolation as indicated.

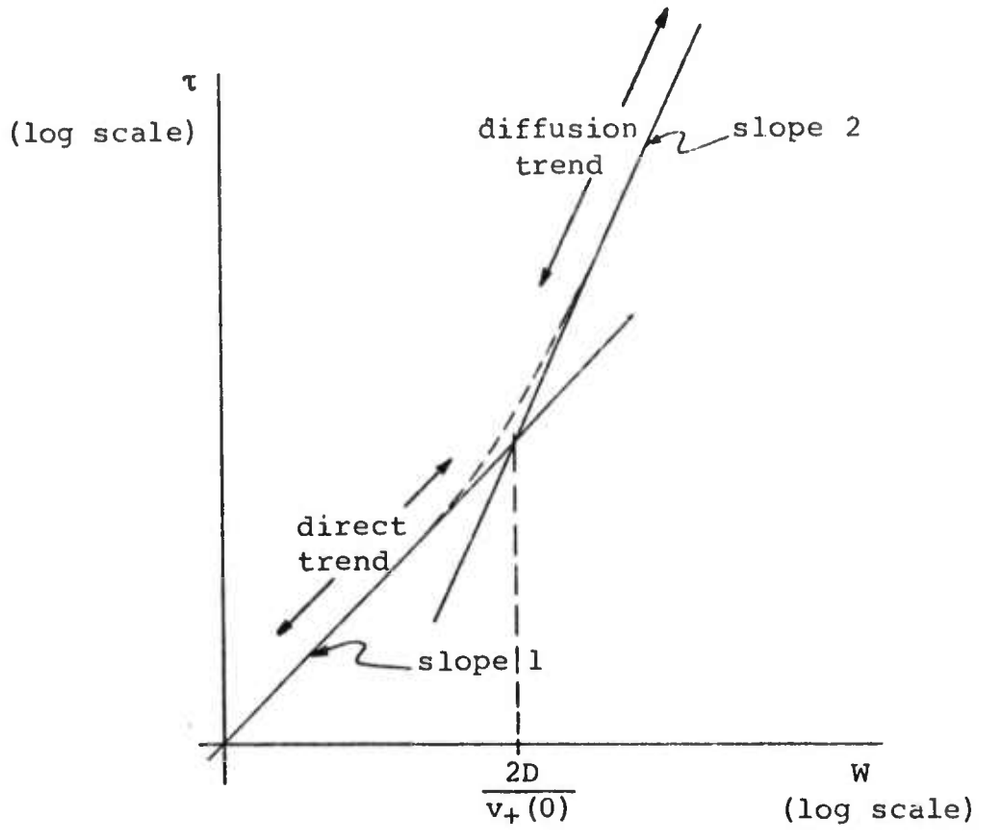


Figure 3 Transit time defined in Eq. (18), showing asymptotes corresponding to flow by direct transition and to flow by diffusion.

IV. The Effect of Neutron Radiation on Noise in Field-Effect Transistors
(K. Wang and E. R. Chenette)

INTRODUCTION

The purpose of this report is to present the results of some recent work on the effects of neutron radiation on the noise-performance of various types of field-effect transistors.

Problems of the effects of radiation on the performance of semiconductor devices have been investigated by many workers [Ref. 1-7]. Shedd, Buchanan, and Dolan have investigated radiation effects on JFET's and reported that JFET's show superior tolerance to permanent damage by neutron radiation. Some of the JFET's used in our studies are of the same radiation-resistant types studied by these workers.

METHOD

The purpose of the work was to determine the effect of neutron irradiation on the noise-performance of selected devices. Most emphasis was given to the units provided by W. M. Shedd of the Air Force Cambridge Research Laboratory. These units (RAD 102, RAD 201, and RAD 201-1B) are typical of those studied by Shedd, Buchanan, and Dolan [Ref. 2]. They are devices designed for optimization of the neutron radiation tolerance.

Other types of devices studied included Texas Instruments type SFB 8558 (now the 2N6450) very low-noise JFET's and some specially-designed MOSFET's as well as commercially available MOSFET's.

The characteristics and noise performance of the devices were determined both before and after being irradiated in the University of Florida Training Reactor.

For most of the work it has been found adequate to restrict the noise measurements to the low-frequency region and to use the University of Florida Real Time Spectrum Analyzer to determine the equivalent noise resistance in the frequency range from 3.15 Hz to 80 kHz.

Measurements have been made of the spectra with temperature as a parameter. The temperature of the device under test could be varied from 77°K to 315°K for these studies. As will be described below these temperature dependent spectra can be used to estimate effective time constants associated with the charge capture processes.

RESULTS

It is important to document as clearly as possible the characteristics of the University of Florida Training Reactor. Fig. 1 and Table 1 can be used to determine the approximate energy distribution of the neutron flux [Ref. 8]. Fig. 1 shows the relative neutron flux as a function of group number. Each group number has associated with it a lethargy, U, or energy, E, range. Lethargy and energy are related by the expression

$$U = \ell n\ (10^7/E).$$

Figs. 2-5 show the static characteristics of several devices before and after the norminal 2.5×10^{14} neutrons/cm^2 irradiation (E>10K eV)(10 minutes at 10 kW). The negligible change is consistent with the results reported by Shadd, Buchanan, and Dolan. This small change is further substantiated by Table II which shows a tabulation of the measured g_m and I_{DSS} of these several devices before and after irratiation.

Fig. 6 shows a comparison of typical spectra of equivalent noise resistance (R_n) before and after irradiation. Several interesting (and perhaps nearly trivial) observations can be made.

(1) The device which suffered least change in its noise perfcrmance was the heavily gold-doped RAD 201-1B. The generation-recombination noise resulting from the large density of gold defects apparently dominated the noise performance even after this level of irradiation.

(2) The device which suffered the greatest change was the SFB 8558. However even after irradiation its noise performance was superior to that of the best "radiation-resistant" device after it had been irradiated.

It seems on the basis of this result that there is sensibly no such thing as a radiation-hardened low noise device. All that is possible is that the noise performance of a device can be degraded before irradiation in such a way that radiation will cause little change.

Fig. 7 shows spectra of R_n over the range from 10 Hz to 80 kHz with temperature as a parameter. The same data are displayed in Fig. 8 as a function of temperature. Figs. 9 and 10 shows the temperature dependence of the dc drain current and the g_m (measured at 1 kHz) for this same device.

Figs 11, 12, 13, and 14 show spectra of several different MOSFET's before and after irradiation for various operating conditions. Fig. 11 shows the effect of irradiation first at 2.5×10^{14} neutron/cm^2 and then at 2.5×10^{15} neutrons/cm^2. The effect of the radiation on the quiescent operating conditions is tabulated on the figure. Figs. 12, 13, and 14 show only the effect of the total 2.5×10^{15} neutron/cm^2 flux accumulation.

DISCUSSION OF RESULTS

The equivalent noise resistance of a JFET with its noise performance dominated by generation-recombination processes with a single time constant is given by the expression [Ref. 9, 10 and 11]

$$R_n = \text{Constant} \cdot \frac{a^3}{Z \cdot L} \cdot \frac{N_T \cdot f_T(1-f_T)\tau_T}{1+\omega^2\tau_T^2} \cdot F(V)$$

Here a, Z, and L are the channel height, width and length; N_T is the generation-recombination trap density; f_T is the probability a trap is occupied; τ_T is the effective time constant associated with the trap; and F(V) is a complex function of the basic voltages.

The geometrical term in the above expression may be manipulated to show that

$$R_n \propto \frac{W_p^{3/2}}{Z\ L\ N_D^{3/2}} \ .$$

Here W_p is the pinch-off voltage and N_D is the density of impurities in the conducting channel. Therefore a device with a high channel doping denisty and with the shallow channel (consistent with a low pinch-off voltage) should yield low noise due to radiation defects. From our experience, the best device having the lowest low-frequency noise is SFB 8558 which is made by Texas Instruments, Inc. and is now carrying the trade name 2N6450. We have measured units with R_n as low as 110 ohms at 3.15 Hz. This unit has a channel doping of about $1.5 \times 10^{16}/cm^3$. The pinched off voltage is only 1.92 volts, but the drain current at zero gate voltage is large (30 mA) which means the channel width is larger than that of other devices. All of them result in better noise performance, both before and after irradiation.

The radiation-resistant RAD 102 and RAD 201 devices also have extremely high channel doping densities. The density for the RAD 102 is about $3 \times 10^{16}/cm^3$ and for the RAD 201 is about $1 \times 10^{17}/cm^3$. Consequently, the RAD 201 after irradiation shows less noise than the RAD 102 after irradiation. The RAD 201-1B had in addition been heavily gold doped to reduce minority carrier lifetime and thus to reduce transient currents resulting from photocurrents. The effect of this heavy gold-doping is clearly seen in the sharply defined g-r spectrum of this device.

It is well known that the temperature dependence of the time constant of the charge fluctuation can be determined by measuring temperature dependence of ideal g-r spectra [Ref. 12]. These data can in turn be used to determine defect energy level and capture cross section.

Unfortunately most of the spectra (Fig. 6, Fig. 7, and Figs. 12-14) don't show a strong dependence on a single g-r time constant. It's more likely that a wide distribution of defect energy levels have been generated by the neutron radiation.

It is interesting to note the frequency dependence of the peaks in the curves of R_n versus temperature as shown on Fig. 8. This information is used to obtain Fig. 15 which can be considered to be an experimentally determined graph of the temperature dependence of an apparent "time constant" for generation-recombination processes to neutron-produced defects.

Much more noise has been given by MOSFET's before and after radiations than by JFET's. Strong 1/f noise has consistently appeared both before and after neutron radiation in the lab-fabricated p-channel MOSFET's, while RCA n-channel MOSFET 3N152 shows interesting $1/f^{1/2}$ noise after neutron radiation.

CONCLUSIONS

Measurements have been made of the effect of neutron radiation, on the noise performance of typical low-noise and radiation hardened JFET's and of some commercially available and lab-fabricated MOSFET's. Radiation levels which had almost no effect on other device parameter produced significant increases in the noise of all devices studied.

It appears that a generation-recombination model with a wide distribution of time constants may provide an adequate description of the problem.

REFERENCES

1. B. Buchanan, R. Dolan, and S. Roosild, "Comparison of the Neuton Radiation Tolerance of Bipolar and Junction Field Effect Transistors," Proc. of IEEE , Vol. 55, pp. 2188-2189, December 1967.

2. W. Shedd, B. Buchanan, and R. Dolan, "Radiation Effects on Junction Field Effect Transistors," IEEE Transaction on Nuclear Science, Vol. NS-16, No. 6, pp. 87-95, December 1969.

3. W. L. George, "Optimization of the Newtron Radiation Tolerance of Junction Field Effect Transistors," IEEE Transaction on Nuclear Science, Vol. NS-16, No. 6, pp. 81-86, December 1969.

4. S. S. Naik, and W. G. Oldnam, "Neutron Radiation Effects in Junction Field-Effect Transistors," IEEE Transaction on Nuclear Science, Vol. NS-18, No. 5, pp. 9-17, October 1971.

5. B. L. Gregory, S. S. Naik, and W. G. Oldham, "Neutron Produced Trapping Centers in Junction Field Effect Transistors," IEEE Transaction on Nuclear Science, Vol. NS-18, No. 6, pp. 50-59, December 1971.

6. H. E. Kern, "Noise from Neutron Induced Defects in Junction Field Effect Transistors," IEEE Transactions on Nuclear Science, Vol. NS-17, No. 6, pp. 256-261, December 1970.

7. I. N. Krishnan, and T. M. Chen, "The Effects of Electron Bombardment on the Noise in Junction Gate Field Effect Transistors," To be published in Solid State Electronics.

8. C. A. Thompson, "Neutron Flux Calculations for a Graphite Moderated Twenty Per Cent Enriched Reacter," Master Thesis, University of Florida, Gainesville, Florida, January 1961.

9. C. T. Sah, "Theory of Low-Frequency Generation Noise in Junction Gate Field-Effect Transistors," Proc. of IEEE, Vol. 52, pp. 795-814, July 1964.

10. P. O. Lauritzen, "Low-Frequency Generation Noise in Junction Field-Effect Transistors" Solid State Electronics, Vol. 8, pp. 41-58, January 1965.

11. A. van der Ziel, "Noise in Solid State Devices and Lasers," Proc. of IEEE, Vol. 58, pp. 1178-1206, August 1970.

12. J. W. Haslett, and E. J. M. Kendall, "Temperature Dependence of Low-Frequency Excess Noise in Junction-Gate FET's," IEEE Transaction on Electron Devices, Vol. ED-19, August 1972.

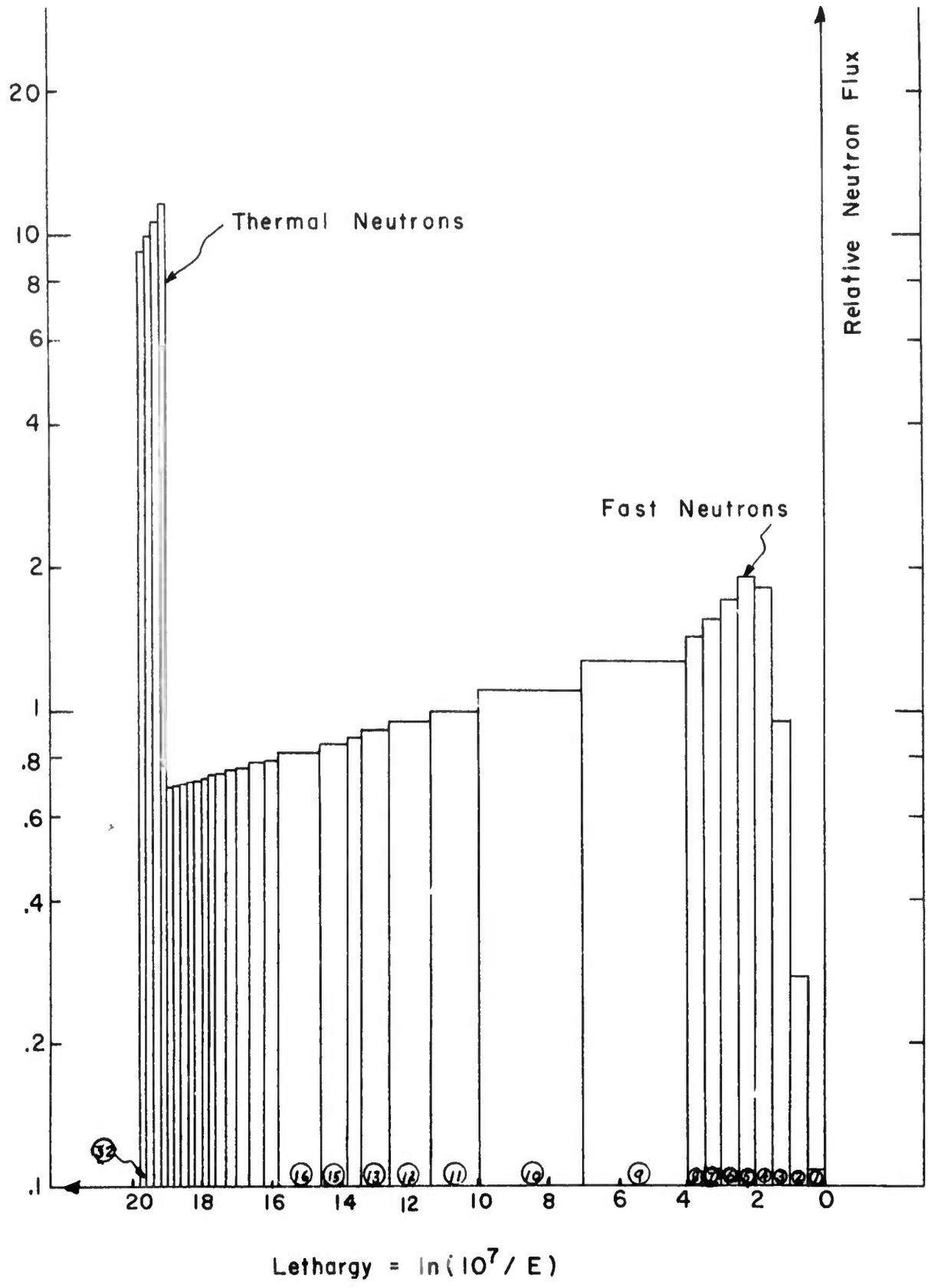
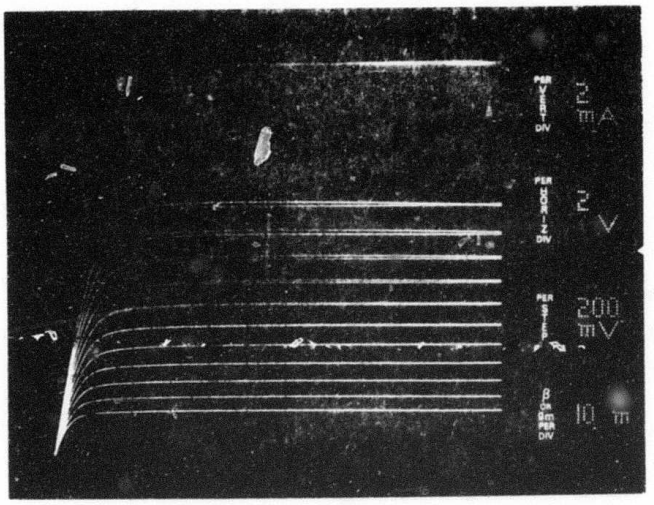


Fig. 1 Relative neutron flux as a function of lethargy.

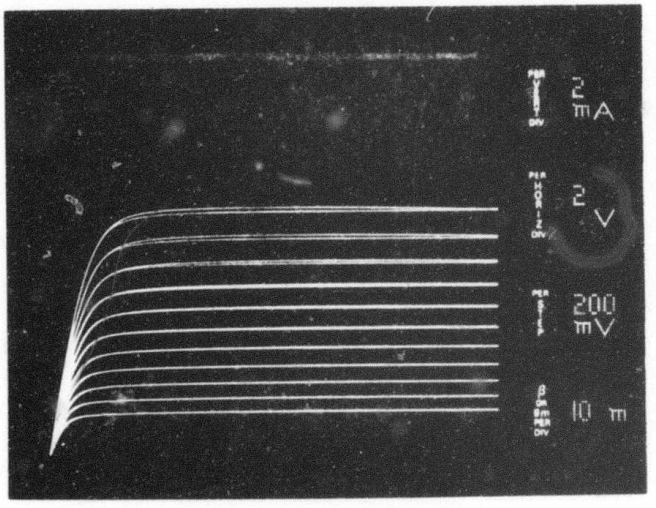
MAGNUM MULTIGROUP ENERGY INTERVALS FOR 32 GROUPS

Group	Lethargy Limits	Energy Limits	Group	Lethargy Limits	Energy Limits
1	0. -	10^7 ev-	17	15.8-	1.375ev-
2	0.5-	6.065×10^6-	18	16.2-	0.9214-
3	1.0-	3.679×10^6-	19	16.6-	0.6176-
4	1.5-	2.231×10^6-	20	17.0-	0.4140-
5	2.0-	1.353×10^6-	21	17.3-	0.3075-
6	2.5-	8.208×10^5-	22	17.6-	0.2272-
7	3.0-	4.98×10^5-	23	17.8-	0.1860-
8	3.5-	3.02×10^5-	24	18.0-	0.15230
9	4.0-	1.832×10^5	25	18.2-	0.1247-
10	7.0-	9.118×10^3-	26	18.4-	0.1021-
11	10.0-	4.54×10^2-	27	18.6-	0.08358-
12	11.4-	112-	28	18.8-	0.06843-
13	12.6-	33.72-	29	19.0	0.05603-
14	13.4-	15.15-	30	19.2-	0.04587-
15	13.8-	10.16-	31	19.4-	0.03756-
16	14.6-	4.564-	32	19.6-19.8	0.03975-0.02518

Table I: Lethargy and energy limits for lethargy groups shown in Fig. 1.

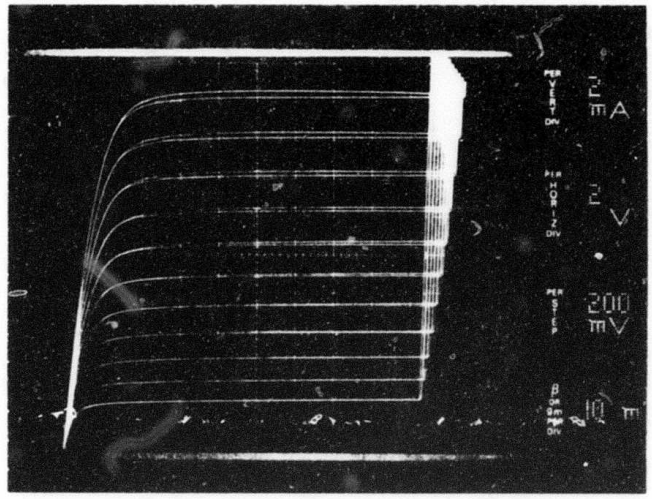


(a)

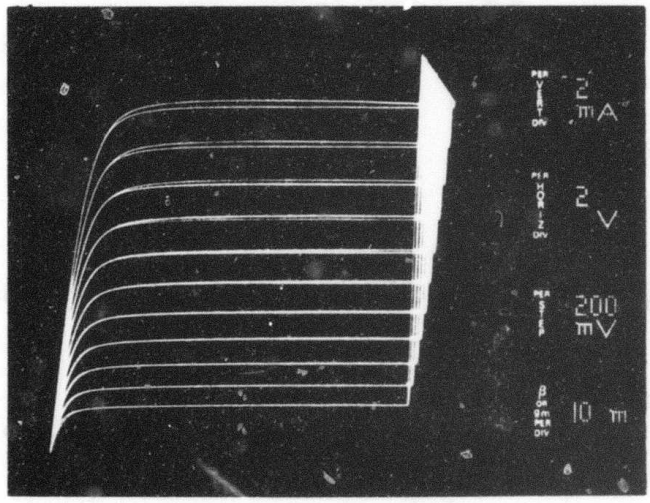


(b)

Fig. 2 DC characteristics for RAD 102 No. 10 (a) Before neutron irradiation. (b) After 2.5×10^{14} neutrons/cm^2 irradiation.

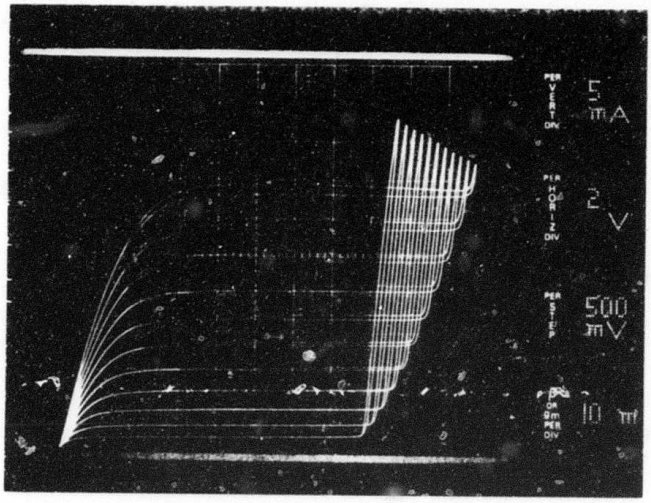


(a)

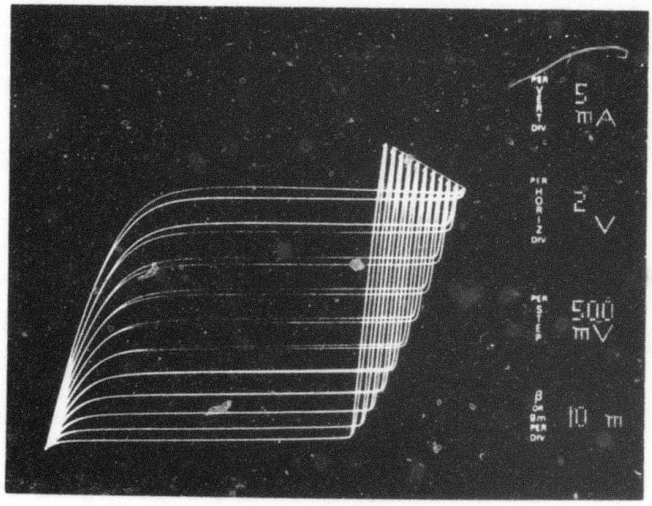


(b)

Fig. 3 D.C. characteristics for RAD 201 No. 19 (a) Before neutron irradiation. (b) After 2.5×10^{14} neutrons/cm^2 irradiation.

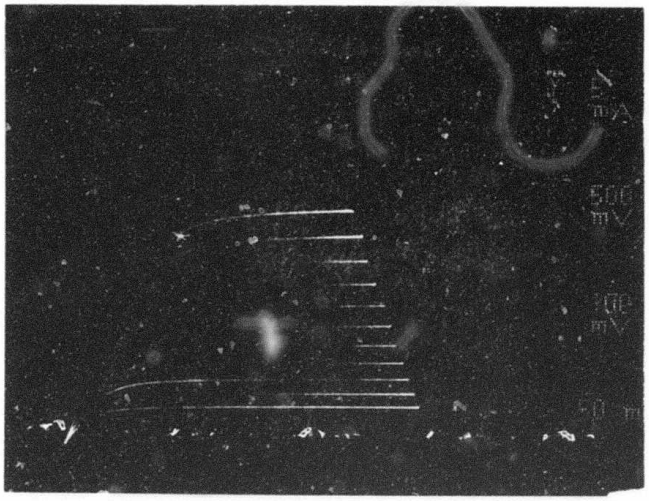


(a)

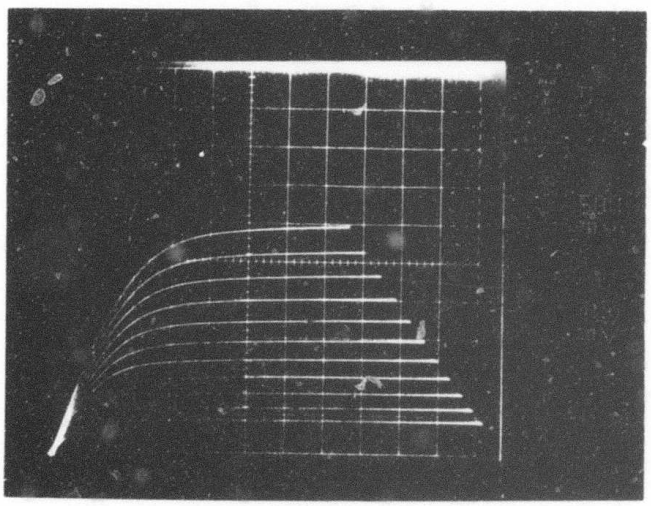


(b)

Fig. 4 D. C. characteristics for gold-doped RAD 201-1B No. 89. (a) Before neutron irradiation. (b) After 2.5×10^{14} neutrons/cm^2 irradiation.



(a)



(b)

Fig. 5 D. C. Characteristics for SFB 8558 No. 16
(a) Before neutron irradiation.
(b) After 2.5×10^{14} neutrons/cm^2 irradiation.

Type	RAD 102	RAD 201	RAD 201-1B	SFB 8558
No.	10	19	89	16
V_{GS} (Volts)	0	0	0	0
V_{DS} (Volts)	5	5	8	5
$I_{DSS}(\Phi=0)$ (mA)	13.0	18.7	34.5	32.3
$I_{DSS}(\Phi=2.5\times10^{14})$ (mA)	12.4	18.2	34.4	30.1
g_m ($\Phi=0$) (v)	7.17×10^{-3}	10.8×10^{-3}	9.11×10^{-3}	34.1×10^{-3}
g_m ($\Phi=2.5\times10^{14}$) (v)	6.90×10^{-3}	10.5x10-3	9.09×10^{-3}	33.2×10^{-3}

Table II: Change of I_{DSS} and g_m with irradiation for JFET's.

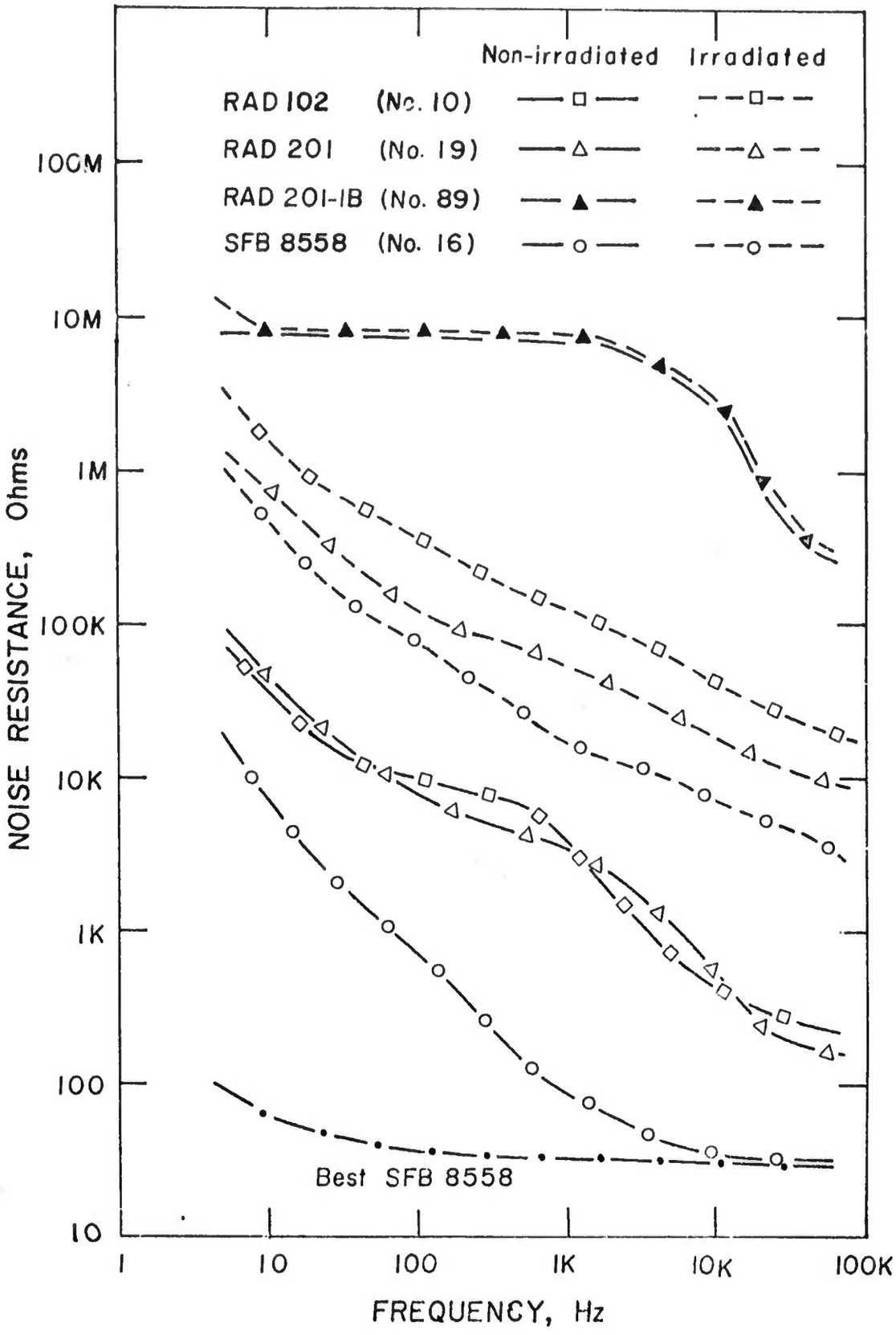


Fig. 6 Noise spectra due to 2.5×10^{14} neutrons/cm^2 irradiation for JFET's.

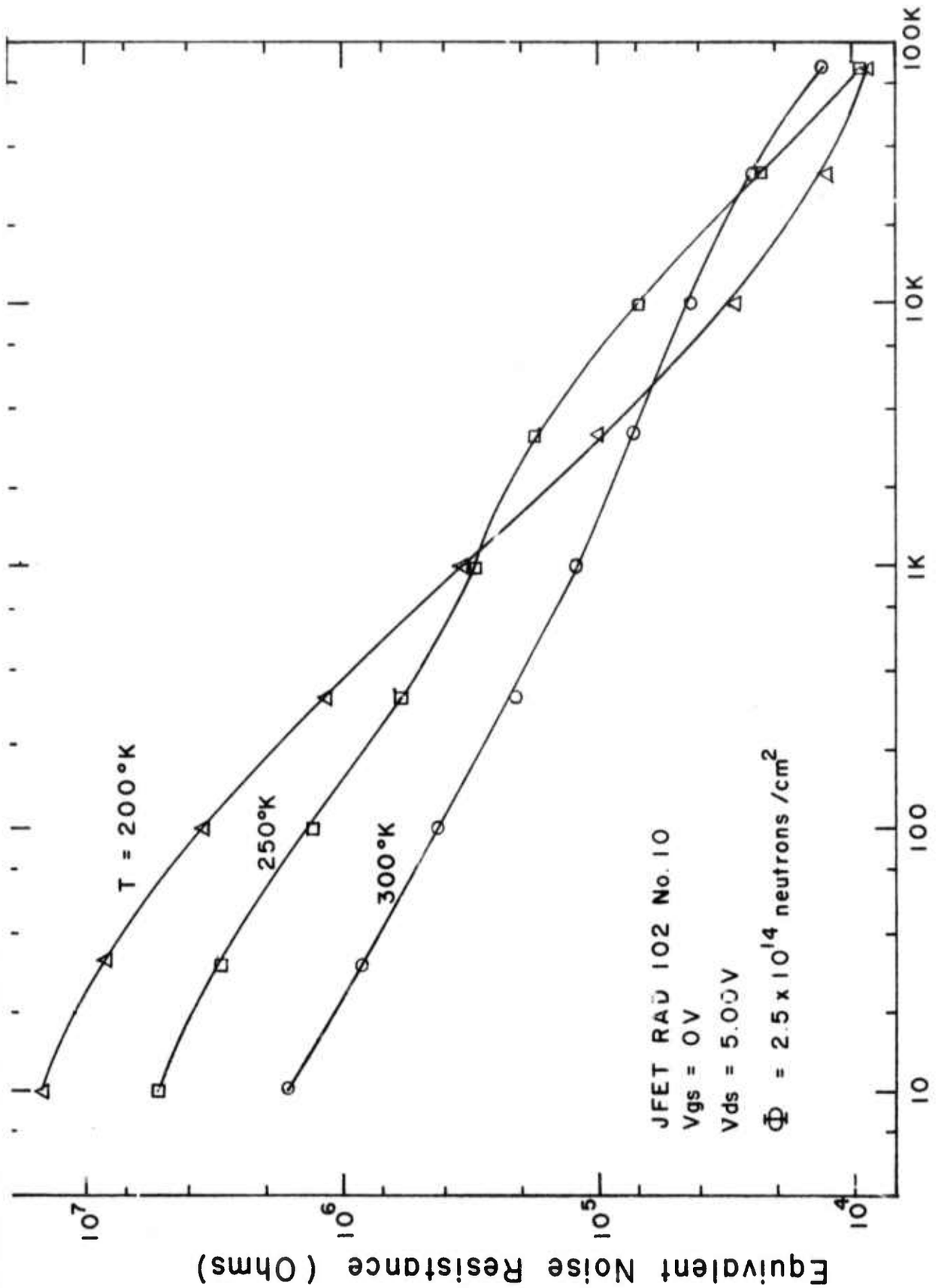


Fig. 7 Noise spectra at different temperatures.

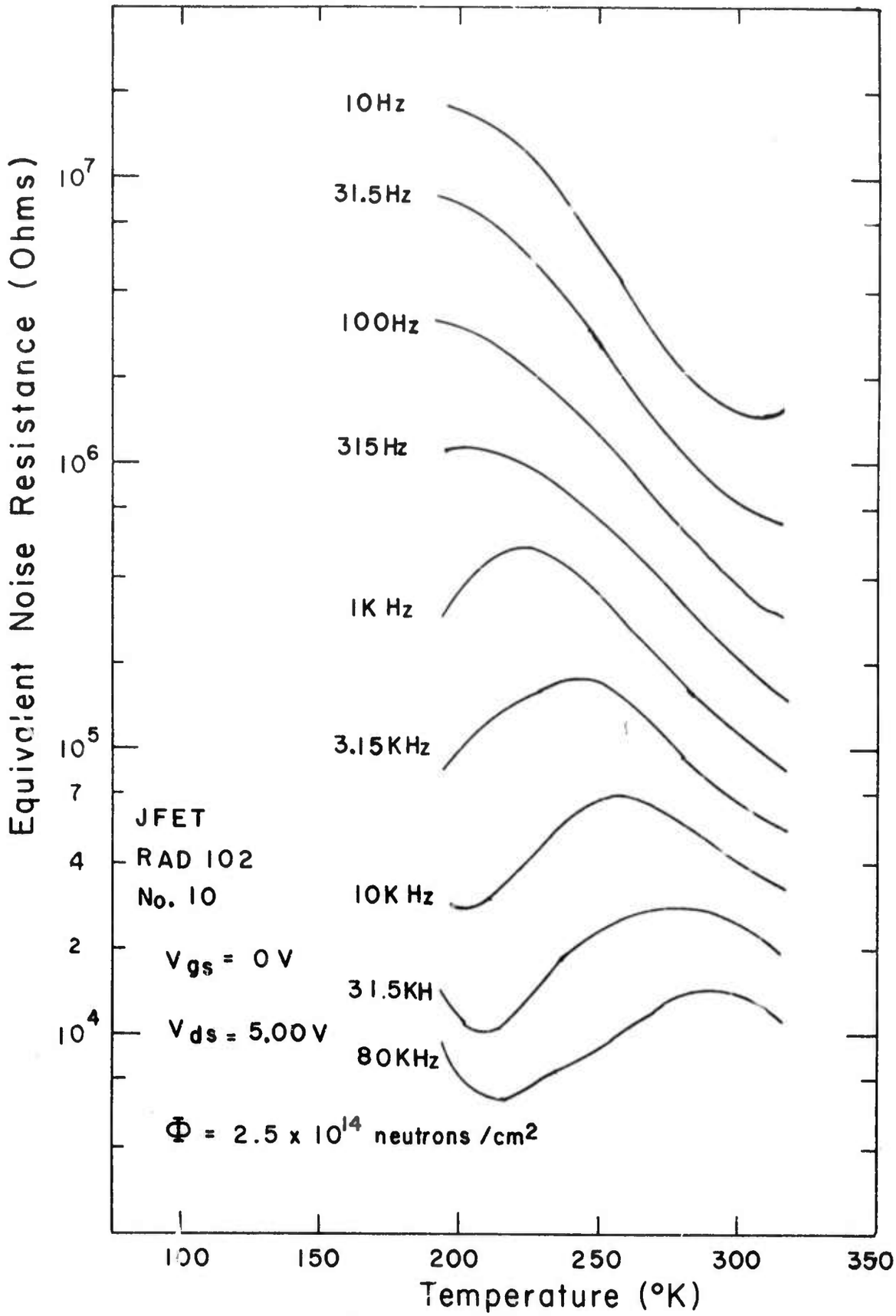


Fig. 8 Noise vs. temperature.

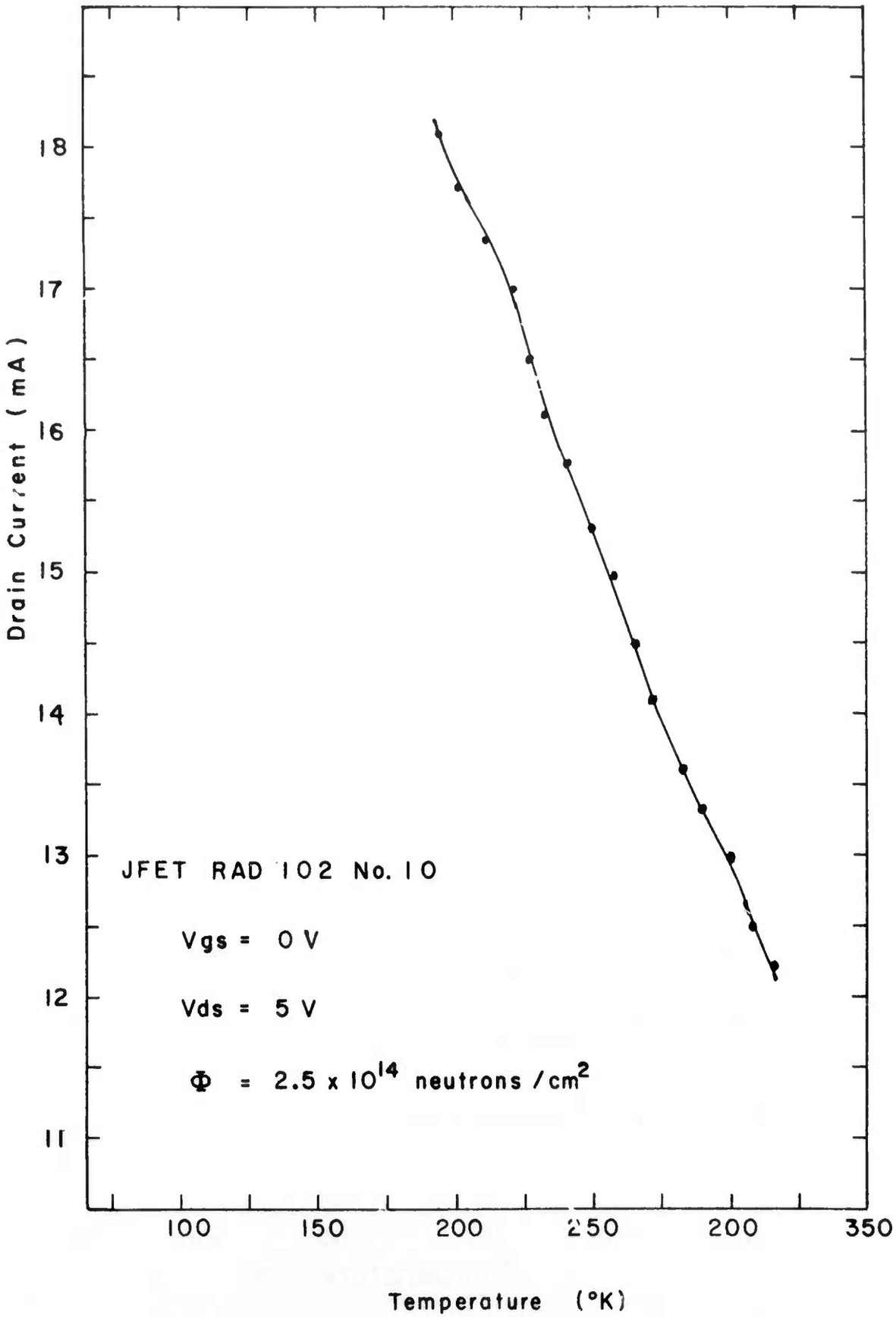


Fig. 9 Drain current vs. temperature.

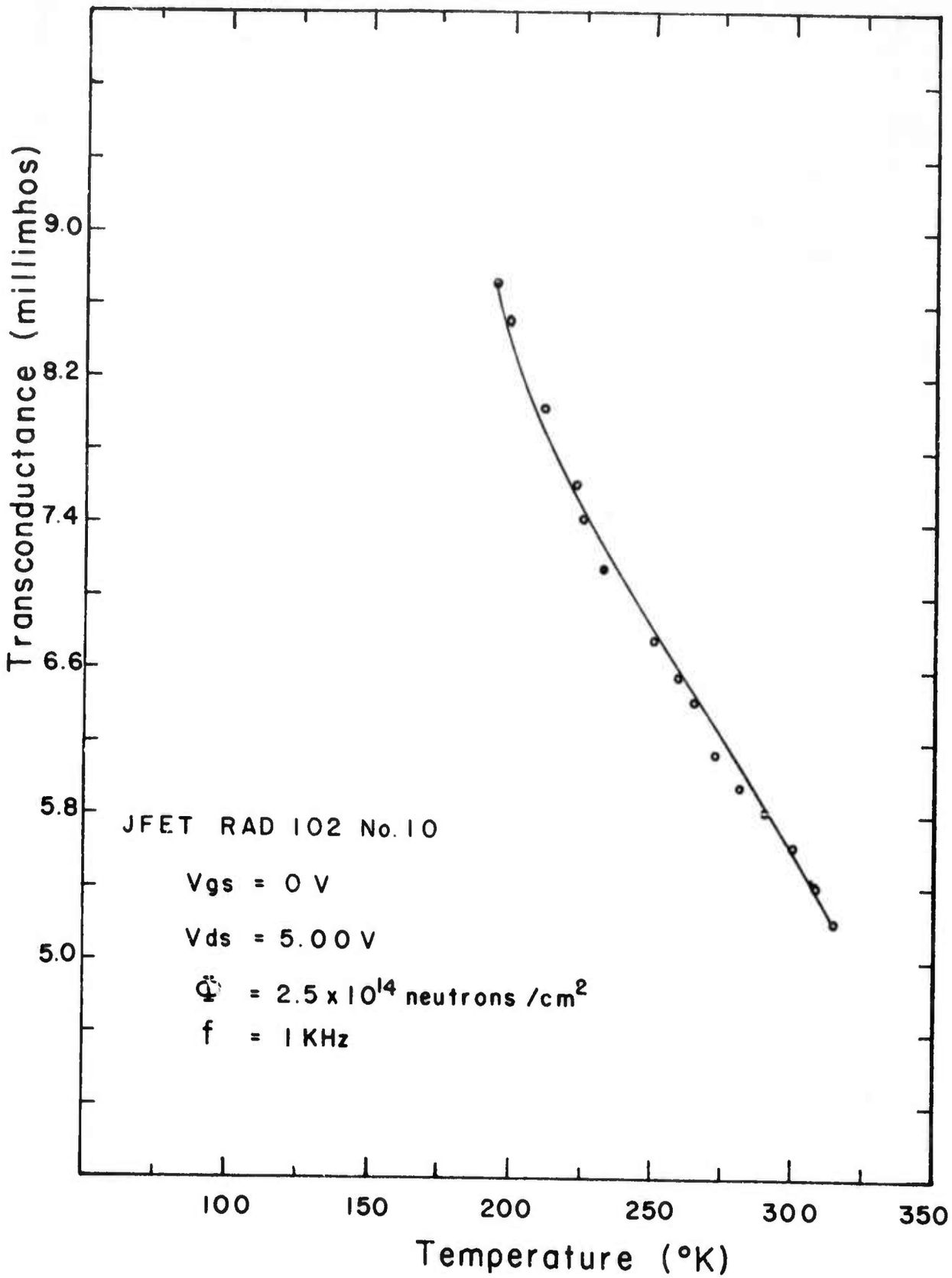


Fig. 10 Transconductance vs. temperature.

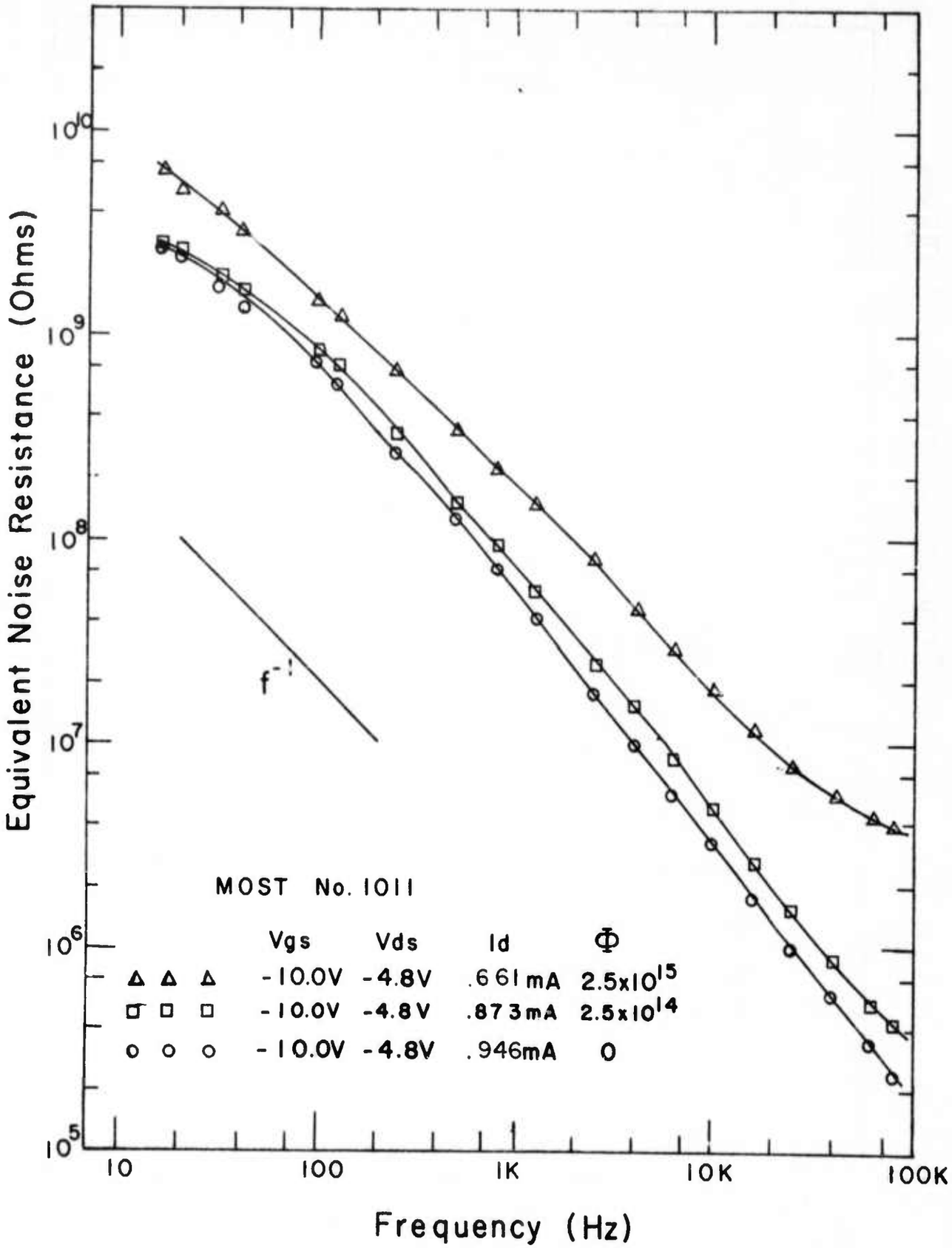


Fig. 11 Noise spectra due to different dose of neutron irradiation on lab-fabrication MOSFET.

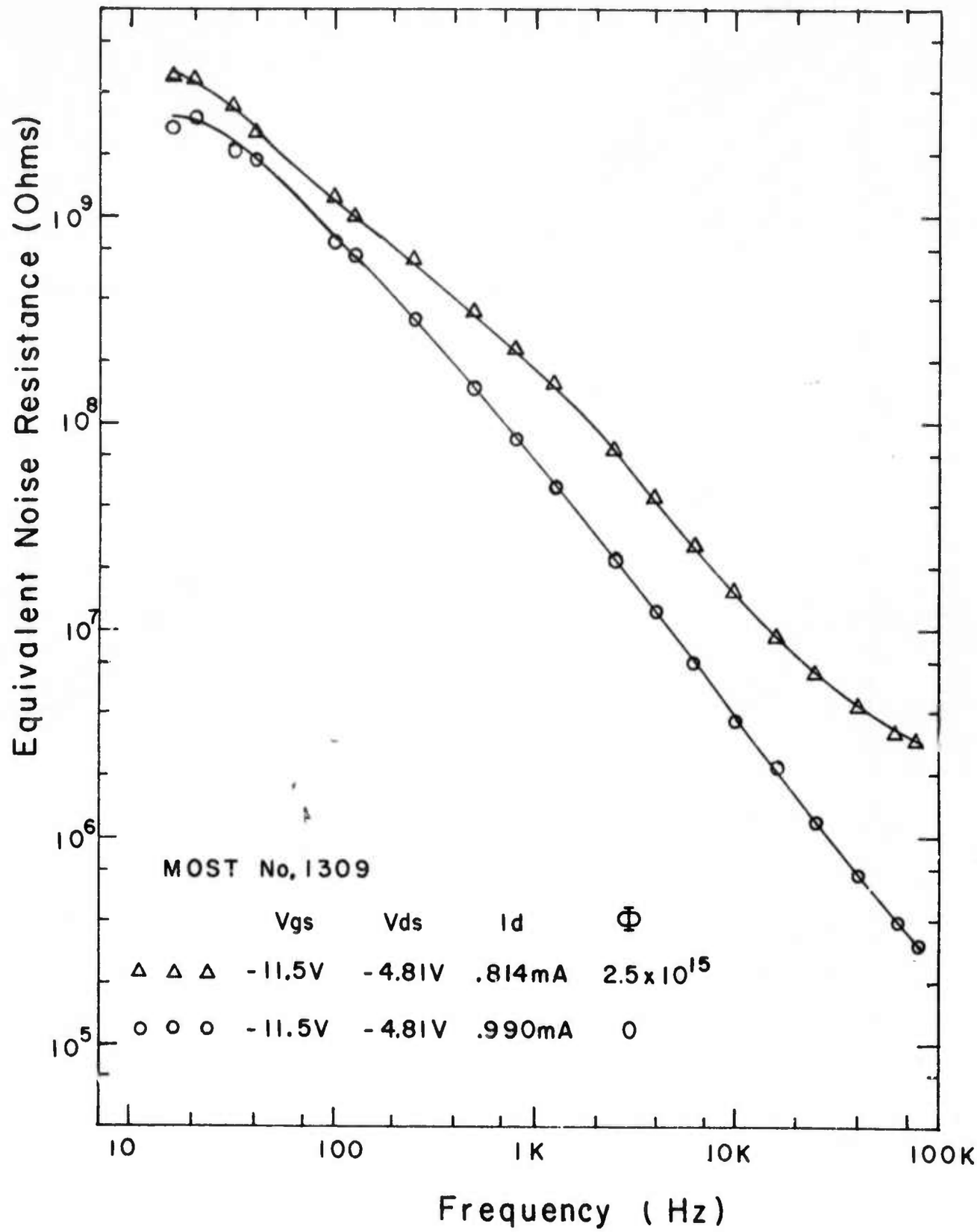


Fig. 12 Noise spectra due to neutron irradiation on lab-fabricated MOSFET.

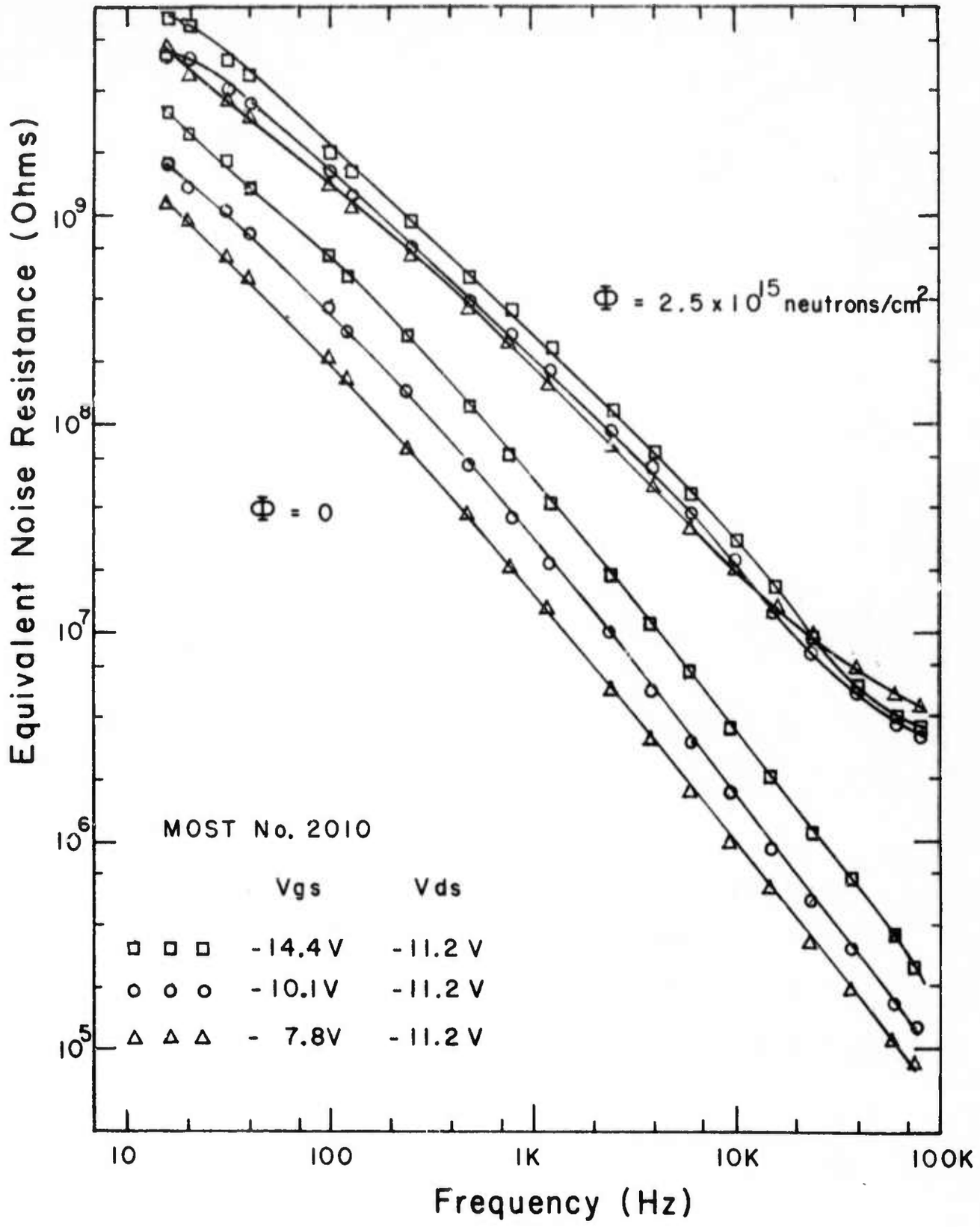


Fig. 13 Noise spectra with gate voltage as a parameter.

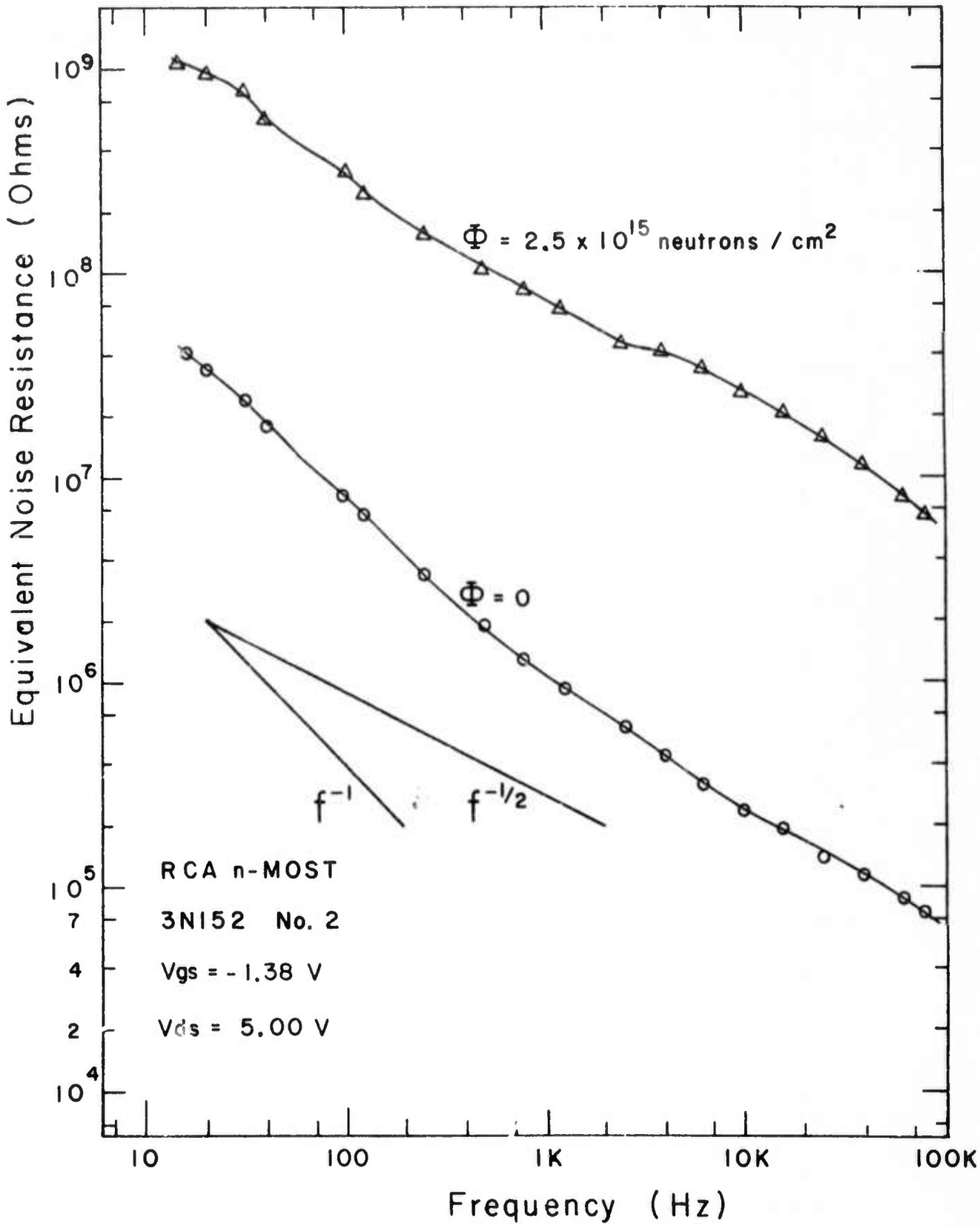


Fig 14 Noise spectra due to neutron irradiation in a conventional MOSFET.

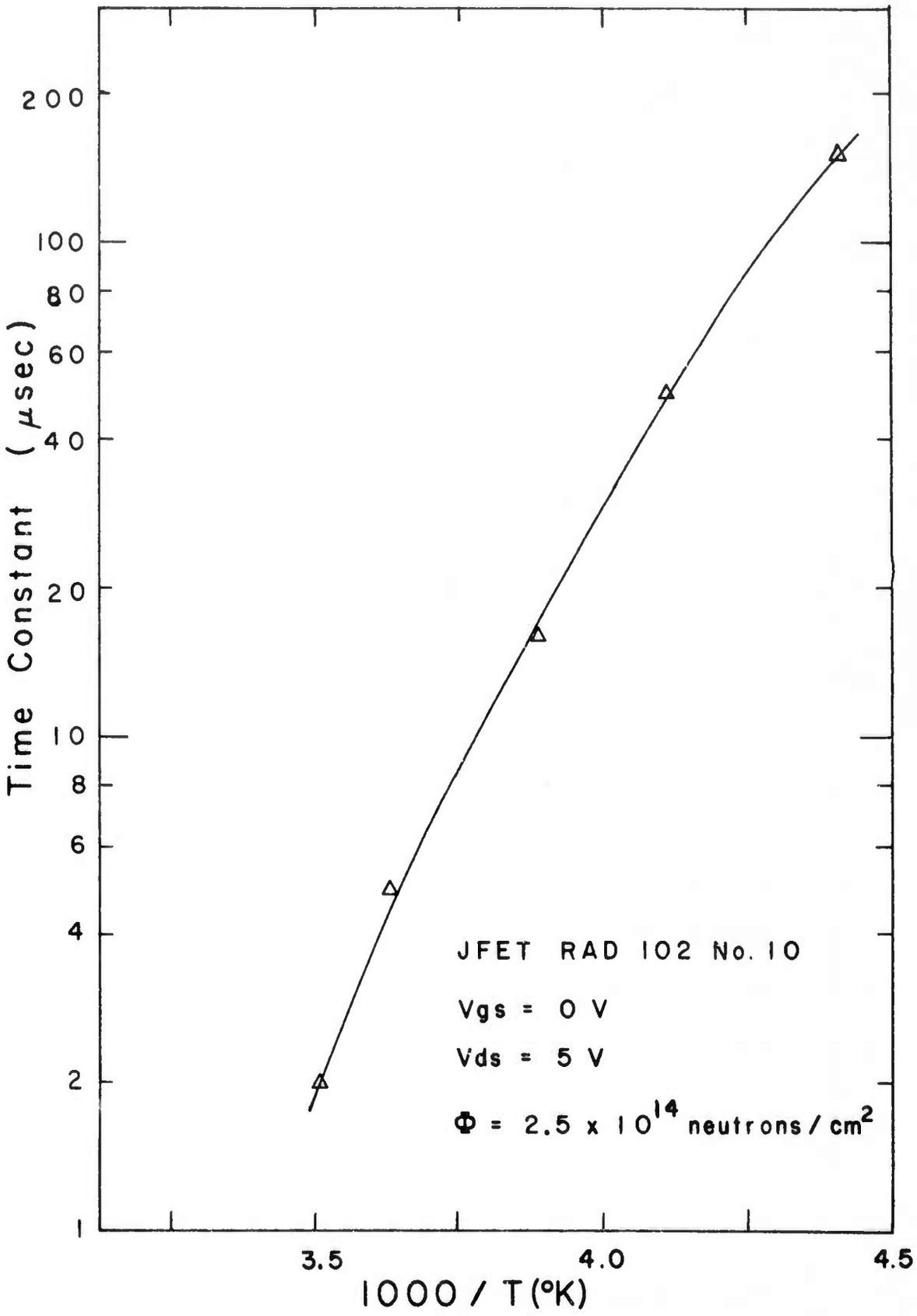


Fig. 15 Characteristic time constant of charge fluctuation at neutron-produced defects vs. temperature.

V. A New High Sensitivity Integrated Silicon Schottky-Barrier Phototransistor (S. S. Li and E. R. Cox, Jr.)

I. Introduction

Phototransistor was first suggested by Shockley, Sparks, and Teal[1] as a variation on the "hook" transistor. In the past few years much interest has been shown in the phototransistor as an image sensing device because of its low cost, low power consumption and long life. With the advent of GaAs laser diodes and other III - V LED's which radiate light in the visible to near infrared range, the silicon phototransistor has acquired its added significance.

The main advantage of a phototransistor over a Schottky-barrier photodiode is that the former provides with internal current gain via transistor action. However, slow response speed is the main drawback for a phototransistor, as compared to a Schottky-barrier photodiode.

If the incident photons could be efficiently coupled into the depletion region of a Schottky-barrier photodiode, this would result in a high speed and high efficiency detector[2-3]. The conventional technique to achieve this is by depositing a thin metal film (~100 Å) on a semiconductor substrate as proposed originally by Schneider[4]. Li and Wang[5-6] have recently reported a grating type silicon Schottky-photodiode which shows an overall improvement in responsivity and quantum yield as compared with the thin film Schottky-barrier photodiode in the spectral range between 0.4 μm to 1.0 μm.

The Schottky-barrier photodiode is expected to have a better response in the short wavelength side of the spectrum as compared to the phototransistor. However, a combination of these two devices into a single structure should provide an overall improvement in responsivity, quantum yield and response speed. In this paper we propose a new grating type integrated silicon Schottky-barrier phototransistor. The device is constructed by depositing a grating type aluminum film as a Schottky-barrier contact on the collector-base region of the ordinary phototransistor. The objective for the propose device is

to improve the overall responsivity of the conventional phototransistor by the addition of a grating type Schottky-barrier contact on the collector-base region of the phototransistor. The total photocurrent is produced by the Schottky-diode in parallel with the base-collector junction and amplifying by transistor current gain β. Thus, it is expected that the resulting device should have a much higher responsivity and quantum yield than that of the conventional phototransistor.

In this work, the equivalent circuit model, the dc I - V characteristics, the spectral dependence of the responsivity and quantum yield, and the response speed are discussed and analyzed for both the grating type Schottky-barrier phototransistors and the corresponding ordinary phototransistors.

II. General Analysis

2.1 Equivalent Circuit Model

Analysis of the basic operation principles and characteristics of the phototransistor has been given by previous investigators. Gary and Linvill[7] have derived analytical expressions for the quantum efficiency of a photodiode and developed a model for optical phenomena in diodes and transistors; Joy and Linvill[8] have analyzed phototransistor operation in the charge storage mode;Schuldt and Kruse[9] have treated the problem of image resolution of a single phototransistor illuminated by nonuniform light. A more comprehensive numerical analysis of the quantum efficiency as well as noise characteristics for an ordinary silicon phototransistor has been given recently by De La Moneda[10]. Li et al.[11] have reported the computation of quantum efficiency for a metal-semiconductor Schottky-barrier photodiode taking into account the inversion layer effect.

The equivalent circuit model used in the analysis of an ordinary phototransistor is the hybrid-pi model[12], as shown in Fig 1.a. When illuminated, the current generated in the collector-base junction is the sum of the dark leakage current (I_{CEO}) and the photoinduced current (I_{phcb}). The photo-induced current is normally much greater than the

dark current. Under illuminated conditions, the collector current flow is given by (neglecting dark current):

$$I_c = (h_{fe} + 1)\, I_{phcb} \tag{1}$$

where h_{fe} is the forward current gain of the transistor and I_{phcb} is the photoinduced current. This results in a current generator I_{phcb} added to the basic hybrid-pi model connected from collector to base (see Fig. 1.a). If a Schottky-barrier contact is made on to the collector-base region of an ordinary phototransistor, the current generated in the Schottky diode will add to that generated by the collector-base junction. When light impinges upon such device the photocurrent generated in the Schottky contact will also add to the photocurrent generated in the collector-base junction. The sum of these two currents will be the photoinduced base current of the grating structure Schottky-barrier phototransistor. Thus, an equivalent circuit model for such device can be represented by Fig. 1.b. The total collector current for such device is thus given by

$$I_c = (I_{phcb} + I_{phsb})\, (h_{fe} + 1) \tag{2}$$

where I_{phsb} denotes the photoinduced current due to the Schottky contact. This results in an additional current generator I_{phsb} in parallel with I_{phcb}. An additional capacitance is also introduced as a result of the Schottky contact and is labelled as C_{sb} in the model.

2.2 Responsivity and Quantum Yield

The penetration depth of the incident photons is inversely related to the absorption coefficient. For silicon, the absorption coefficient exceeds 10^3 cm^{-1} for $\lambda \leq 0.8$ μm[13]. Thus, the short wavelength photons usually are absorbed near the surface of the device. The addition of the grating Schottky-contact to the surface of the ordinary phototransistor will greatly increase the photoresponse of the device. This is indeed the case, as will be shown later.

The responsivity of a photodetector is defined as the photocurrent per incident photon power intensity. This can be expressed by

$$R = \frac{I_{ph}}{P_{in}} \tag{3}$$

where I_{ph} is the measured photocurrent in ampere and P_{in} is the power intensity of the incoming light in watt.

The overall quantum yield of a phototransistor is defined by

$$\eta_T = \left(\frac{I_{ph}}{q}\right)\left(\frac{h\nu}{P_{in}}\right) = \left(\frac{I_{ph}}{P_{in}}\right)_{\lambda} \times 1.24 \tag{4}$$

where q is the electronic charge; hν is the incident phonton energy, and λ is the wavelength of incident photon in μm. For an ordinary phototransistor, the overall quantum yield η_T in Eq. (4) is related to the quantum yield of the collector-base junction, η_{cb}, by the expression[10]:

$$\eta_T = (h_{fe} + 1)\eta_{cb} \tag{5}$$

Numerical computation of η_{cb} for an ordinary silicon phototransistor has been given by De La Moneda[10] for different junction depth surface recombination speed and carrier lifetimes. From Eqs. (4) and (5), it is noted that by measuring η_T and computing η_{cb} one can estimate the current gain, β, for the phototransistor.

For a grating-type Schottky-barrier phototransistor, the overall quantum yield can be written as:

$$\eta'_T = (h_{fe} + 1)\ (\eta_{cb} + \eta_{sb}) \tag{6}$$

where η_{sb} is the quantum yield of the Schottky-barrier contact, the general expression for this is given by Li et.al.[11]

The spectral dependence of the measured responsivity and the quantum yield for the grating type Schottky-barrier phototransistor and the corresponding ordinary phototransistor will be presented in a later section.

2.3 Response Speed

The response speed for a deplete-mode photodiode is the sum of the carrier transit time in the depletion layer region and the RC

time constant of the photodiode. The RC time constant is the product of the series resistance of the diode and the junction capacitance. If a load is attached, the total series resistance will be the sum of the series resistance of the device plus the load resistance. Thus, the response speed of the detector can be expressed approximately by:

$$t_r = \frac{W}{v_s} + R_s C_s \qquad (7)$$

where R_s is the total series resistance, C_s is the junction capacitance, W the depletion layer width and v_s the scattering limited velocity.

III. Device Fabrication

Design considerations of a grating-type silicon Schottky-barrier photodiode have been discussed in details by Wang and Li[6]. For devices reported in this paper, n-type silicon wafers of 12 ~ 20 Ω-cm (<110>orientation>) were used. The grating spacing of Schottky contact is chosen to be 12.5 μm. With this arrangement a ten-volt bias is required in order to completely deplete the region underneath and between the strips of the aluminum film. The geometry of our grating type silicon Schottky-barrier phototransistor phototransistor and its dimensions are shown in Fig. 2. A schematic illustration of a grating type silicon Schottky-barrier phototransistor is also shown in Fig. 3. The structure,geometry and size of the ordinary phototransistor is similar to that of Fig. 2 with the exception that no aluminum grating contact is made between collector-base region.

We shall next briefly describe the fabrication processes for device structure shown in Fig. 2.

3.1 Cleaning and Oxidation

The silicon wafer was first rinsed with deionized water and immersed in hot nitric acid for 5 minutes. This was followed by a one-second dip on HF and a deionized water rinse. Immediately after cleaning, the wafer was placed in the oxidation furnace at 1150°C and under wet oxygen atmosphere for 15 minutes.

3.2 Base and Emitter Diffusion

After using photolithograph process, the wafer was dipped in dilute HF prior to the boron predeposition to eliminate surface contamination. The wafer was then placed in the predeposition oven at 1008°C for 8 minutes and drive-in furnace at 1150°C for 20 minutes for base-diffusion. The emitter diffusion process was performed by passing phosphine gas in the oven for 6 minutes and the drive-in was done at 1050°C. In both the base and emitter diffusion, precautions have been taken in each fabrication step so that high quality phototransistors can be obtained.

3.3 Contact Mask, Evaporation, Alloging and Wiring

By employing photolithograph method, windows were opened in the oxide so that ohmic and Schottky-barrier contact can be made in the device. Aluminum deposition was done in a vacuum evaporator with a vacuum better than 10^{-6} torr. Immediately after evaporation the grating pattern was made by employing the photolithographic process and the device was then placed in the alloying furnace at 480°C for 15 minutes.

The device mounting was accomplished by scribing the wafer with a diamond cutter and the chips mounted on standard three terminal packages. Thin gold leads were attached to the contact by using thermal compression bonder.

VI. Experimental Results and Discussions

4.1 Leakage Current (I_{CEO}) Measurements

In a phototransistor, the reverse bias leakage current provides a good measure of the quality of the collector-base junction. For the case of grating structure Schottky-barrier phototransistor, the leakage current is expected to be higher due to the additional thermionic emission current contributed from the Schottky contact and the surface leakage current.

Fig. 4 shows the I - V characteristics for several grating structure phototransistors (P - 10 to P - 13) and the corresponding ordinary phototransistors (0 - 3 and 0 - 5) under reverse bias conditions. Note that the leakage current for our ordinary phototransistor is found to be

less than 1nA for $V_{CE} \leq 30V$, while the leakage current for our grating-type Schottky-barrier phototransistor is almost one order of magnitude greater than that of our ordinary phototransistors. This may be due to the increase in surface leakage current in the grating contact and longer thermionic emission current due to the lower barrier height for Aℓ-n Si Schottky-contact[14].

The leakage current for an ordinary phototransistor is a result of the generation-recombination current of the collector-base junction multiplied by the forward current gain of the phototransistor. For the grating structure device, the leakage current is the sum of the generation-recombination current of the collector-base junction and the thermionic emission current plus surface leakage current of Schottky contact multiplied by the forward current gain of the phototransistor. To increase the signal-noise ratio of a phototransistor, the leakage current should keep as low as possible.

4.2 Responsivity and Quantum Yield Measurements

Fig. 5 shows the spectral dependence of the responsivity (R, μA/μW) for several grating structure phototransistors (P-10, P-12, and P-13) and the corresponding ordinary phototransistors (O-1, O-5, and O-6). The results show that a substantial gain (about 4 times) in responsivity is obtained over the spectral range from 0.4 μm to 1.0 μm for our P-series device as compared to our O-series device. For example, the responsivity for device P-12 is 4.082 μA/μW at 0.9 μm and 10.06 μA/μW at 0.6328 μm; the responsivity for device, O-5, constructed at the same time under identical conditions, is 1.429 μA/μW at 0.9 μm and 2.619 μA/μW at 0.6328 μm.

The substantial increase in responsivity for our P-series device can be attributed to the following facts: (1) the Schottky-barrier contact is built on the surface of collector-base region of the transistor where most of the incident photons are absorbed in and adjacent to the depletion region of the Schottky contact and thus produce useful photocurrent, (2) the grating structure reduces the reflection loss of the ordinary metal film and provides a more effective way of coupling the incident light into the photoactive region of the device; and (3) high quantum yield.

The spectral dependence of the quantum yield for our P-series device and our O-series device is calculated from Eq. (4), the measured photocurrent and the power intensity of incident light. The result is displayed in Fig. 5 for devices P-10, P-12 and P-13 as well as devices O-1, O-5 and O-6.

The numerical data for the collector-base quantum yield of the ordinary silicon phototransistor, η_{cb}, as a function of photon wavelength has been given by De La Moneda[10]. By using these values for η_{cb} and Eq. (5), the forward current gain is found to be around 12, which is slightly higher than expected. This is due to the fact that the measured photocurrent is much greater than the dark current in most of our devices. As the photocurrent increases the beta of the transistor will also increase from the value found under dark conditions.

The incident light intensity was measured by using both optical power meter (silicon detector with sensitivity down to 10^{-10} W) and a RCA S-20 photomultiplier. Readings from these instruments were compared and an average value taken.

It is worth mentioned that measurement of the photocurrent versus applied reverse bias voltage, V_{CE}, has also been made for both P-series and O-series devices. The results show that the photocurrent depends slightly on the reverse bias voltage for $V_{CE} \geq 2V$.

4.3 Response Speed and Bandwidth

Since the grating structure Schottky-barrier phototransistor operates as a depletion mode photodetector, the lower limit of the response speed is equal to the carrier transit time across the depletion region. However, since the RC time constant of the device is ususally longer than the carrier transit time, the response speed is usually limited by the RC time constant of the detector system. Our P-series device, operating at V_{CE} = 10V, has a series resistance around 85 Ω and a capacitance of about 4 pf, corresponding to a response speed of about 0.34 nsec (see Eq. (7)).

To determine the bandwidth of our P-series devices experimentally, we have made use of the photomixing technique[15]. The beat frequency between the adjacent laser modes of a gas laser is given by[16]

$$\Delta\nu = \nu_{n+1} - \nu_n = \frac{c}{2d} \tag{8}$$

where c is the speed of light in free space, d is the distance between the two reflectors of the laser cavity, and n is the laser mode index. In our experiment, we employed a He-Ne gas laser that has a 2-mW output power at $\lambda = 0.6328$ μm, and d = 27 cm. Thus, this laser may produce beat frequencies at 560 MHz and 1.12 GHz.

The above experiment was done by using the He-Ne laser as a light source, our P-series device as an optical mixer, and an H-P 851 B spectrum analyzer to display the detected optical signal from laser. The result is shown in Fig. 7 for a 560 MHz optical beat signal from He-Ne laser detected by our P-series device. This result leads us to the conclusion that our P-series device has a bandwidth equal or greater than 560 MHz.

V. Conclusions

Several grating-type Aℓ-nSi Schottky-barrier phototransistors have been fabricated and examined in this work. It was found that the responsivity and quantum yield for our grating structure phototransistor was greatly improved as compared to the corresponding ordinary phototransistor in the entire spectrum measured (0.4 ~ 1.0 μm). The leakage current of our P-series device is found to be much greater than our O-series device. This discrepancy could probably be corrected if Pt is used instead of Aℓ for Schottky contact. The responsivity of the Schottky-barrier phototransistor could be further improved by increasing the photoactive area of the Schottky contact. Some of the important parameters for our P-series device are summarized in Table 1.

Table 1. Summary of Some Important Measured and Computed Parameters for a typical grating-type Aℓ-nSi Schottky Phototransistor

Collector doping concentration:	$N_C = 2.34 \times 10^{14}$ cm^{-3}
Epitaxial thickness:	8.9 μm
Device photoactive area:	8.24×10^{-4} cm^2
Schottky contact area:	3.6×10^{-4} cm^2
Spacing between Aℓ-gratings:	12.5 μm
Series resistance:	85 Ω
Capacitance at 10V:	4 pf
Responsivity at 0.6328 μm:	10.06 μA/μW
SNEP at 0.6328 μm:	1.95×10^{-14} W
D_λ^* at 0.6328 μm:	1.02×10^{12} cm-Hz$^{1/2}$/W
Bandwidth:	$\geq$ 560 MHz

Table 1. Summary of Some Important Measured and Computed Parameters for a typical grating-type Aℓ-nSi Schottky Phototransistor

Collector doping concentration:	$N_C = 2.34 \times 10^{14}$ cm^{-3}
Epitaxial thickness:	8.9 μm
Device photoactive area:	8.24×10^{-4} cm^2
Schottky contact area:	3.6×10^{-4} cm^2
Spacing between Aℓ-gratings:	12.5 μm
Series resistance:	85 Ω
Capacitance at 10V:	4 pf
Responsivity at 0.6328 μm:	10.06 μA/μW
SNEP at 0.6328 μm:	1.95×10^{-14} W
D_λ^* at 0.6328 μm:	1.02×10^{12} cm-Hz$^{1/2}$/W
Bandwidth:	$\geq$ 560 MHz

REFERENCES

1. W. F. Shockley, M. Sparks, and G. K. Teal, "P-n junction Transistors," Phys. Rev., 83, P. 151 (1951)

2. E. Ahlstrom and W. W. Gartner, "Silicon Surface-Barrier Photocells," J. Appl. Phys., Vol. 33, pp. 2602-2606, (1962)

3. A. J. Tuzzolino, E. L. Hubbard, M. A. Perkins, and C. Y. Fan, "Photoeffects in Silicon Surface-Barrier Photodiodes," J. Appl. Phys., 33, pp. 148-155 (1962)

4. M. V. Schneider, "Schottky-Barrier Photodiode With Antireflection Coating," Bell Sys. Tech. J., 45, P. 1611-1638, (1966)

5. S. S. Li and C. T. Wang, "A New Grating Type Au-nSi Schottky Barrier Photodiodes for 0.4-1.1 μm Photodetection", Proc. of the 22nd. Electronics Components Conference. p. 105-109 (1972)

6. C. T. Wang and S. S. Li, "A New Grating-Type Gold-N Type Silicon Schottky-Barrier Photodiodes", IEEE Trans. on Electron Devices, ED-20, p. 523-526 (1973)

7. P. A. Gary and J. G. Linvill, "A Planar Silicon Phototransistor with an Optimal Spectral Response for Detecting Printed Material", IEEE Trans. on Elec. Devices, ED-15, P. 30 (1968)

8. R. C. Joy and J. G. Linvill, "Phototransistor Operation in the Charge Storage Mode", IEEE Trans. on Elec. Devices ED-151 P. 237 (1968).

9. S. B. Schuldt and P. W. Kruse, "Optical Gain and Spatial Resolution in a Broad-Area Phototransistor", J. Appl. Phys. 39, 5573 (1968).

10. F. H. De La Moneda, "Noise in Phototransistors", Ph.D dissertation, University of Florida. (1970), unpublished.

11. S. S. Li, F. A. Lindholm and C. T. Wang, "Quantum yield of metal-semiconductor photodiodes". J. Appl. Phys., 43, 4124-4129 (1972)

12. J. Bliss, "Theory and Characteristics of Phototransistors", Mototola Semiconductor Products Application Note, AN-440, (1970)

13. W. C. Dash and R. Newman, Phys. Rev. 99. 1151 (1955)

14. S. M. Sze, "Physics of Semiconductor Devices", John Wiley & Sons, p. 397 (1969)

15. B. Herzog, A. P. Rodgers, J. E. Peterson, M. E. Lasser, G. Lucovsky, and R. B. Emmons, "Detection of Ruby-laser axial mode differences with photodiodes", presented at the Spring Meeting of the Optical Society of America, Washington D. C., (1962).

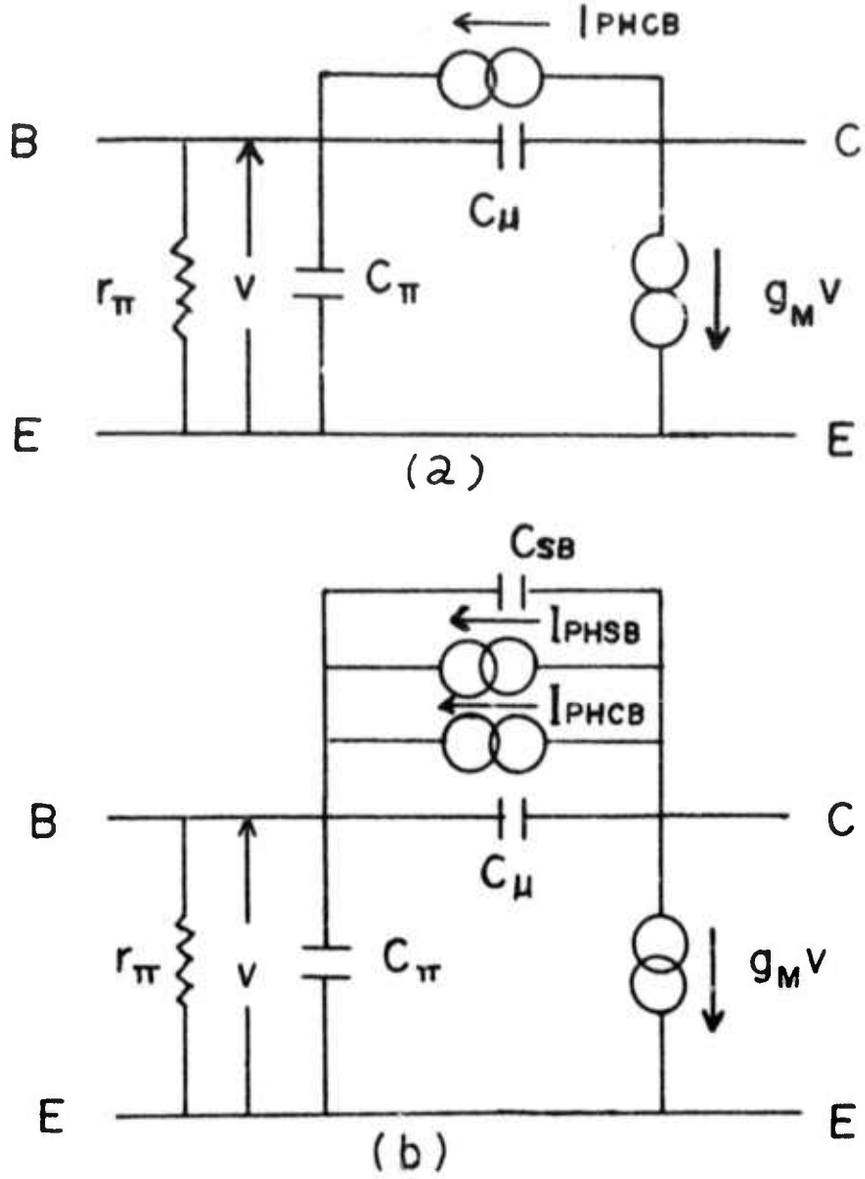


Figure 1 Model for a phototransistor and Schottky-Barrier Phototransistor (a) ordinary phototransistor model (b) grating structure Schottky barrier phototransistor model.

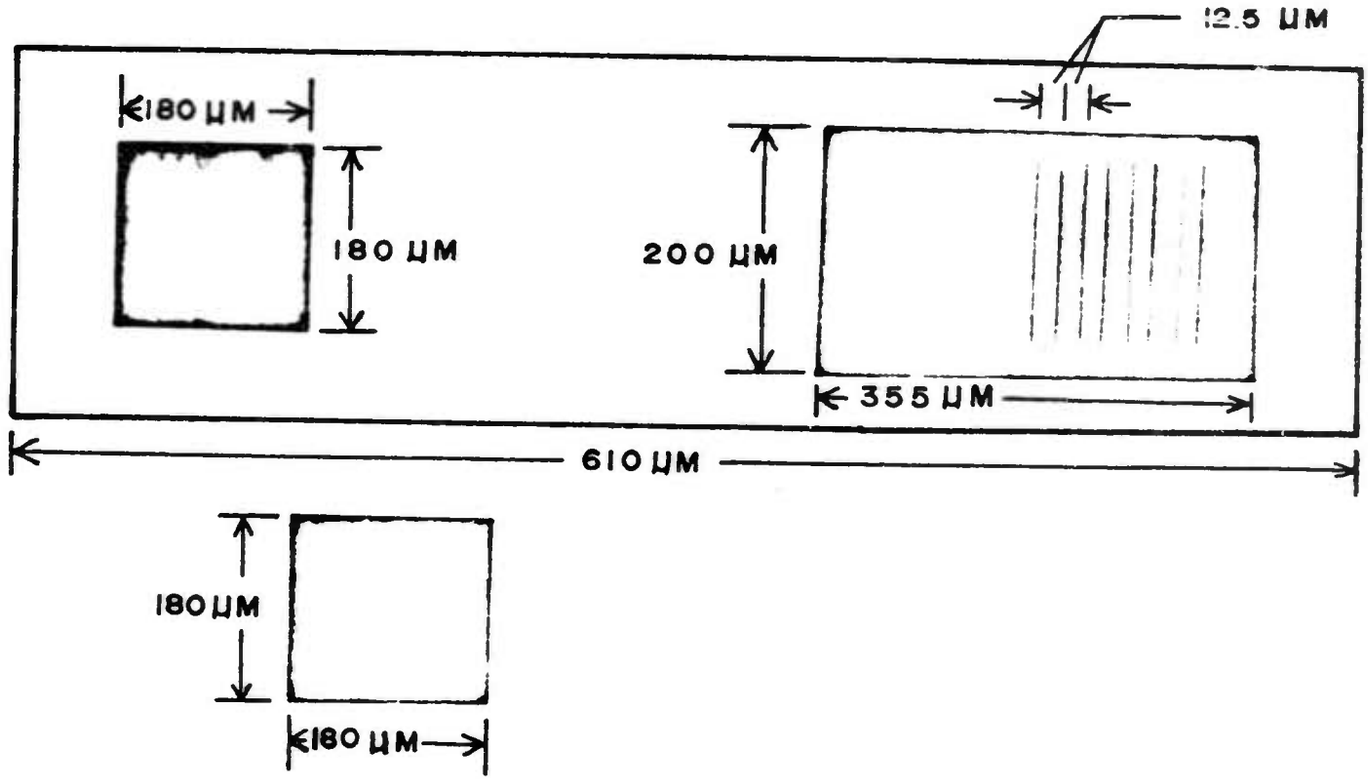


Figure 2 Geometry of a typical grating structure silicon Schottky-barrier phototransistor.

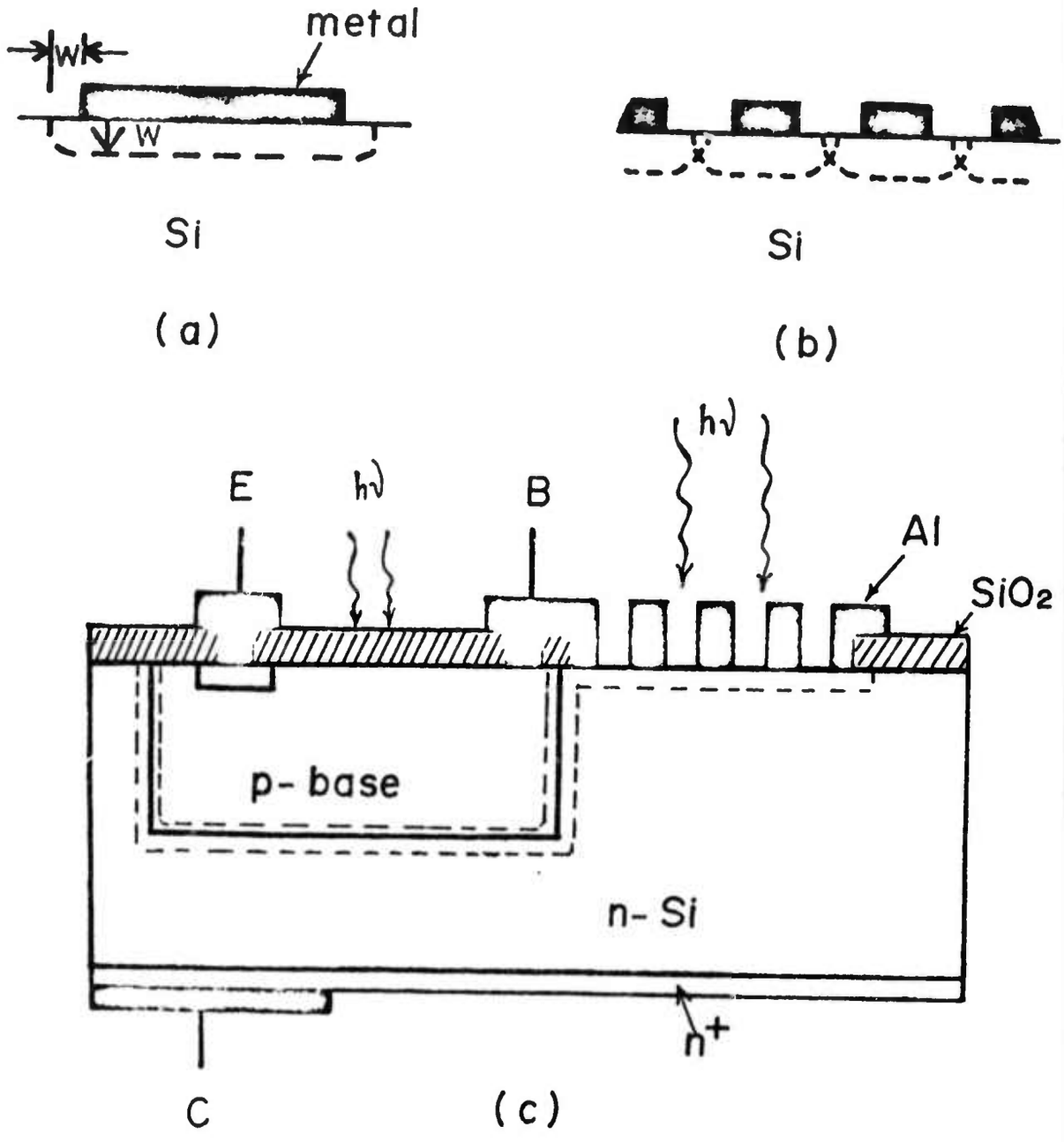


Figure 3 A schematic illustration of a grating structure silicon Schottky-barrier phototransistor.

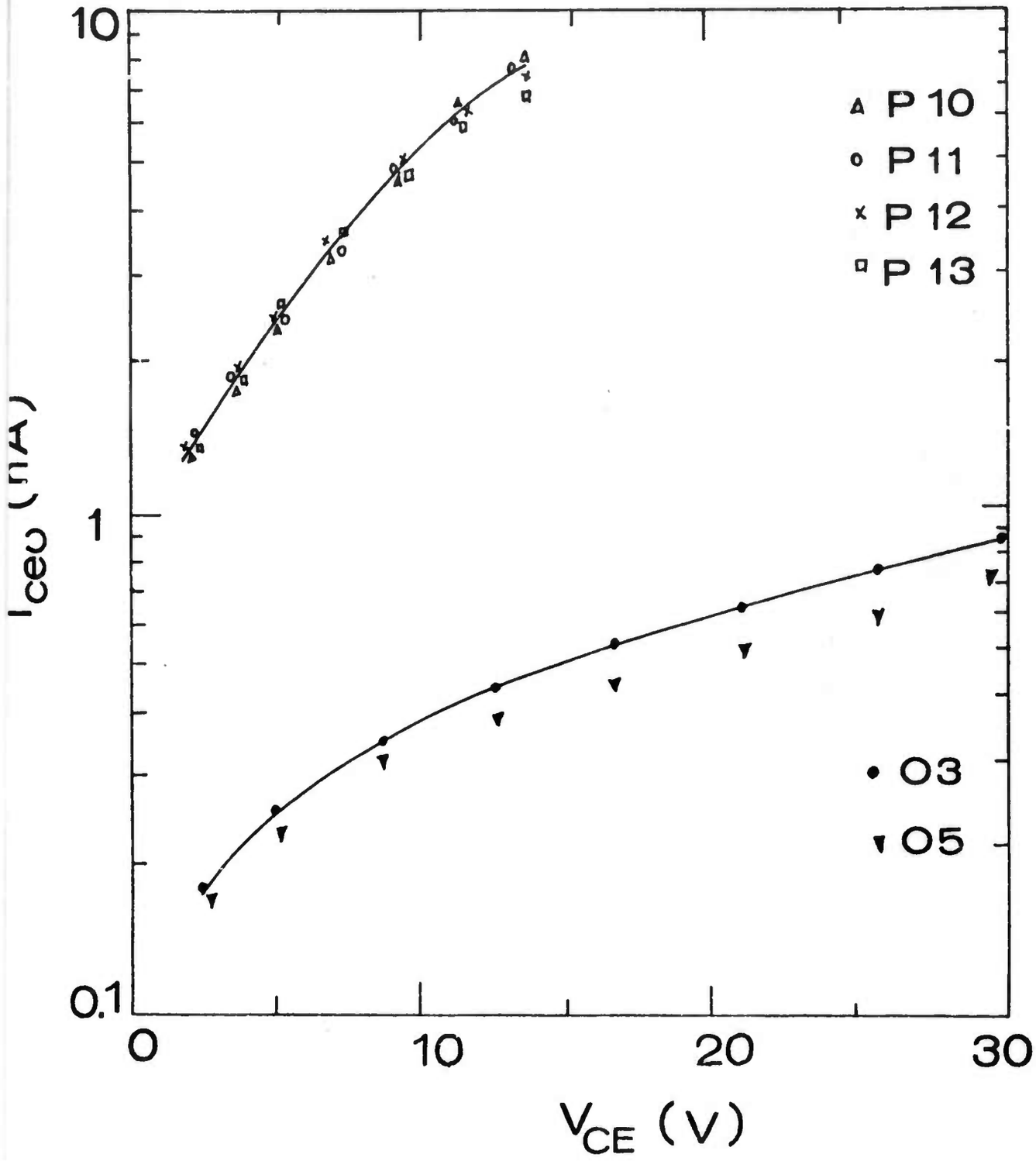


Figure 4 Leakage Current I_{CEO} versus reverse bias voltage, V_{CE} for the grating type Schottky-barrier phototransistors (P-10, P-11, P-12 and P-13) and the ordinary phototransistors (O-3 and O-5).

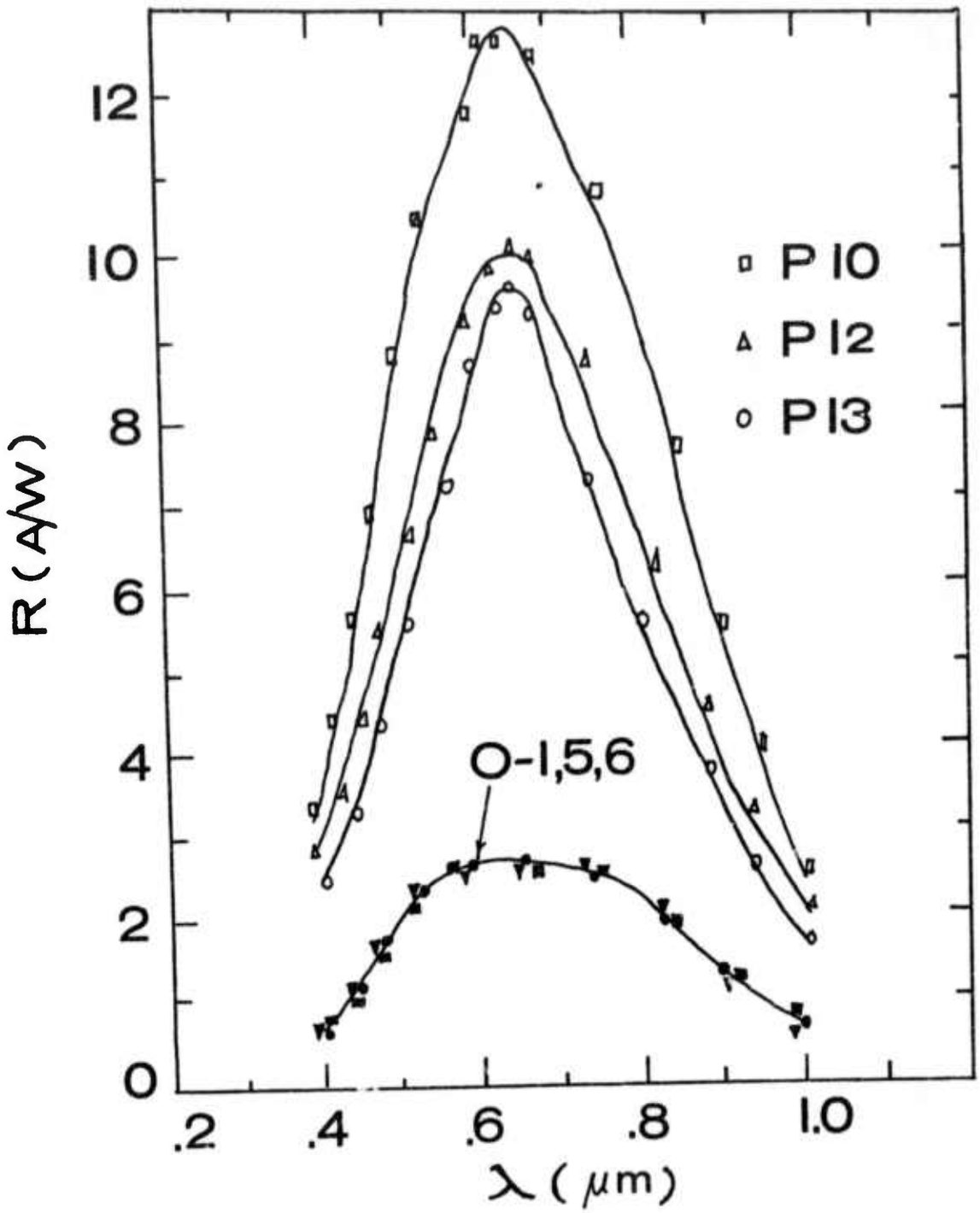


Figure 5 Responsivity versus photon wavelength for the grating type Schottky-barrier phototransistor (P-10, P-12, and P-13) and the ordinary phototransistor (O-1, O-5, and O-6).

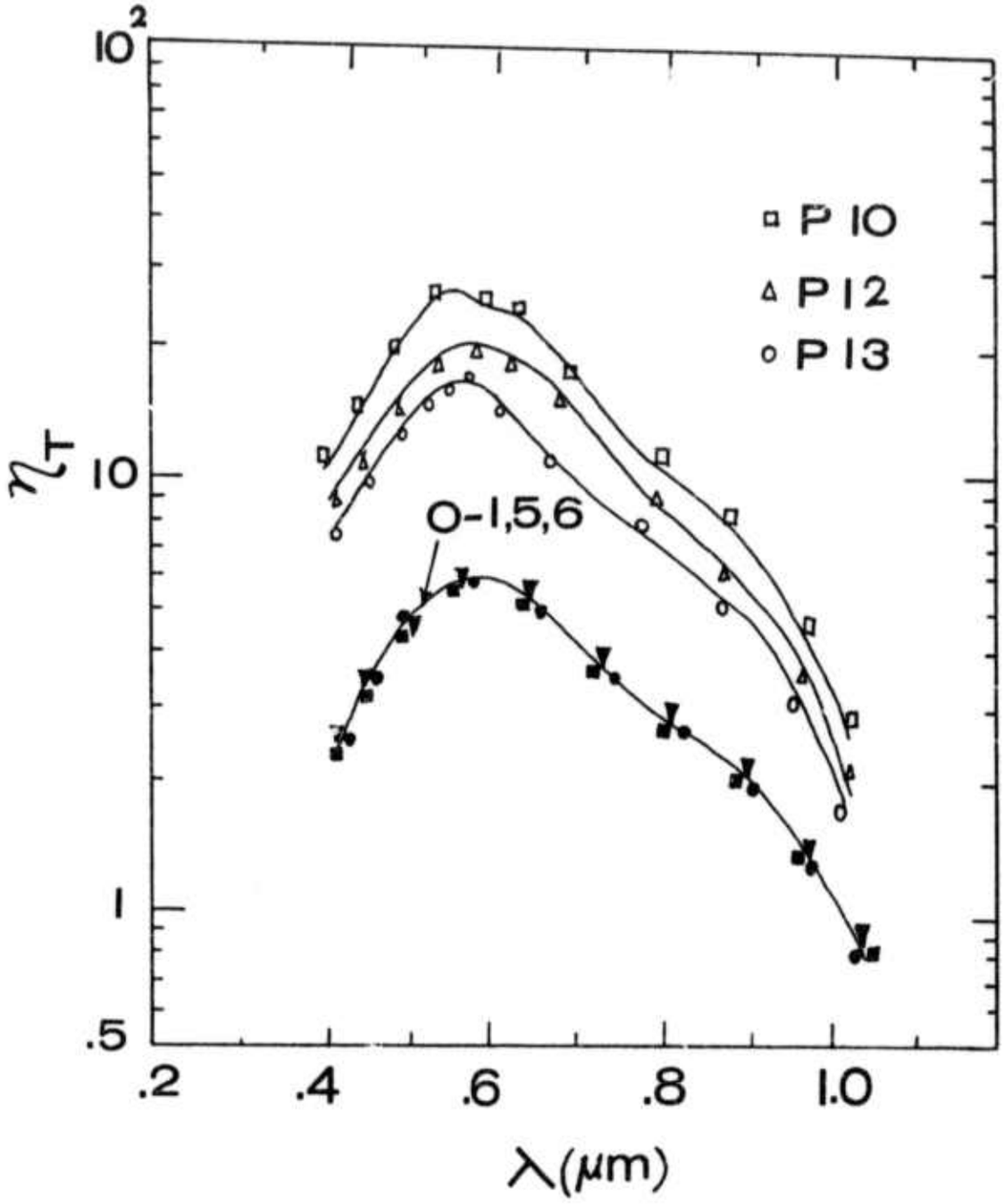


Figure 6 Quantum yield versus photon wavelength for the grating type Schottky-barrier phototransistor (P-10, P-12, and P-13) and the ordinary phototransistor (0-1, 0-5, and 0-6).

Figure 7 Detection of a beat optical signal of 560 MHz from a He-Ne gas laser at λ = 0.6328 μm by device P-12.

VI. Analyses of Transient Capacitance Experiments for Au-GaAs Schottky Barrier Diodes in the Presence of Deep Impurities and the Interfacial Layer (C.I. Huang and S. S. Li)

INTRODUCTION

The effects of deep level impurity and an interfacial layer on the characteristics of a metal-GaAs Schottky barrier diode have been of interest. Goodman[1] presented the first comprehensive descriptions of the metal semiconductor properties in terms of the capacitance-voltage relationship. The interesting transient capacitance behavior in the GaAs Schottky barrier diodes has been observed.[2,3,4,5,6] The transient behavior has been interpreted as a result of the existence of deep level impurity of GaAs which is either intentionally doped or inherited from the crystal growing processes. Although different models have been proposed by these authors in order to interpret the observed capacitance data, the results obtained by them are not in good agreement. For example, the determined thermal activation energy for oxygen impurity in GaAs ranges from 0.57 ev[7] to 0.9 ev[5] from the conduction band edge, comparing with the value of 0.80 ev obtained by optical and Hall effect measurements.[8]

Recently, Sah et al[9] gave a comprehensive and detailed description of the transient capacitance experiments using Au-doped silicon p-n junctions. According to their model the electronic properties of deep impurity centers in semiconductors such as energy level, multiplicity of the charge state, thermal and optical emission rates and capture rates, can be determined readily from the photo-and dark-transient junction current and capacitance measurements. Since their model applies equally well to the metal-semiconductor Schottky diode, we shall in this paper implement their method to analyze our experimental results of the high frequency transient dark-and photo-capacitance measurements of the Au-n type GaAs Schottky barrier diodes in the presence of deep impurities. The effect of an interfacial layer on the diode C-V relation is also considered.

The thermal emission rate of electrons and thermal activation energies of oxygen and chromium impurities in n type GaAs are deduced from these measurements.

II. THEORETICAL BACKGROUND OF TRANSIENT CAPACITANCE EXPERIMENTS

A. Transient Dark-Capacitance Experiment

In this paper, we use a Schottky-barrier diode structure to study the electronic properties of the deep level impurities (i.e., chromium and oxygen) in n type GaAs. To start with, we shall first consider the case of deep acceptor impurity states (e.g., chromium). The deep impurity levels may either be in negative or neutral charge states, depending on the bias condition. If the experiment is performed at a temperature higher than the freeze-out temperature of the deep impurity states, then the deep acceptor impurity states will be in negative charge states at zero bias condition. Now if a large reverse bias voltage V_R is suddenly applied to the Schottky diode, electrons that are originally trapped in the deep acceptor levels within the depletion region of the diode, $n_T(t)$, will be thermally reemitted into the conduction band. It has been shown by Sah et al[9] that

$$n_T(t) = N_T\left\{\left(\frac{e_p}{e_n+e_p}\right) + \left(\frac{e_n}{e_n+e_p}\right)\exp[-(e_n+e_p)t]\right\} \tag{1}$$

where N_T is the deep acceptor impurity density, e_n and e_p are the thermal emission rates of electrons and holes respectively.

The change in the charge density within the depletion region of the diode then results in a change of the depletion layer width and consequently a change of the high frequency capacitance.

The square of the high frequency capacitance as a function of time, t, is given by[9]

$$C^2(t) = \left[\frac{q\epsilon A^2}{2(V_D+V_R)}\right]\left[N_D - N_T \exp(-e_n t)\right] \tag{2}$$

where E is the dielectric constant of the semiconductor; A and V_D are the area and the diffusion potential of the diode respectively; N_D is the shallow donor concentration; the condition $e_n >> e_p$ is also assumed.[5]

The initial value of $C^2(0)$ is obtained from Eq. (2).

$$C^2(t=0) = \left[\frac{q\epsilon A^2}{2(V_D+V_R)}\right](N_D - N_T) \tag{3}$$

and the final value is

$$C^2(t=\infty) = \left[\frac{q\epsilon A^2 N_D}{2(V_D+V_R)}\right] \tag{4}$$

By measuring the time constant, the initial (t=0) and the final (t=∞) values of capacitance, we can determine the values of N_D, N_T and e_n from Eqs. (2), (3) and (4).

If the thermal emission rate of electrons from the deep impurity centers is independent of the electrical field strength, then the activation energy of the deep impurity states can be obtained through the measurement of the temperature dependence of the emission rate by using the relationship[2]

$$e_n = (N_c v_t S) \exp(-E_T/kT) \tag{5}$$

where v_t and S denote the thermal velocity and capture cross section of electrons by deep impurity states respectively. The statistical weighting factor has been assumed equal to unity in Eq. (5).

For the case of donor type deep level impurity (e.g., oxygen in n type GaAs), the time dependence of the square capacitance can be expressed as[9]

$$C^2(t) = \frac{q\epsilon A^2}{2(V_D+V_R)}\left[N_D+N_T(1-\exp(-e_n t))\right] \tag{6}$$

Again, the values of N_D, N_T and e_n can also be determined by Eq. (6) from the measurements of time constant, the initial and final values of capacitance.

B. Transient Photo-Capacitance Experiment

In the transient dark capacitance measurement, the deep acceptor impurity centers were filled with electrons initially at zero bias condition. The filling of electrons at the deep impurity centers can also be achieved by shining the interband ($h\nu > E_g$) light onto the top surface of the device which is reverse biased at a certain voltage. Upon reaching steady state, the recapture of photoinjected electrons by the deep impurities in the

depletion region is balanced by the thermal release of electrons from the impurities.[2] When light is removed, the thermal release of the remaining captured electrons from the deep impurities causes the change in the diode capacitance. From the time constant of this transient capacitance measurement, the thermal emission rate can also be obtained.

III. EXPERIMENTAL DETAILS

A. Preparation of Devices

GaAs wafers were n type oxygen or chromium doped single crystals with faces in the (111) plane. The samples were mechanically lapped and chemically etached in a solution of $3:1:1:H_2SO_4:H_2O_2:H_2O$ at round 90°C. Ohmic contacts were provided on the near surface by evaporation of indium, and alloying at 375°C in hydrogen atmosphere. The front face was chemically polished prior to evaporation of gold dot of area around 3 mm^2. The gold evaporation was performed in vacuum with the background pressure of $\sim 5 \times 10^{-8}$ torr.

The packaging of the device was made by using TO-18 transistor header. By applying silver paste, the ohmic contact side of the diode was "glued" onto the header (collector terminal). After gold wire connections were made, the diodes were baked at 110°C for 24 hours. The basic physical parameters of typical devices fabricated are summarized in Table 1.

B. C-V Measurements

A system for measuring the transient capacitance and transient photo-capacitance was set up. It consists of a Wayne-Kerr B621 Capacitance Bridge, a low noise amplifier, a wave analyzer, a He-Ne gas laser (λ=0.6328 μm) and an X-Y recorder. The system is calibrated such that the deviation from the balanced value of the capacitance bridge, $\Delta C = C(t=\infty)-C(t)$, measured at 100KHz, is linearly proportional to the DC output of the wave analyzer. The procedures for transient dark-capacitance and photo-capacitance are similar to those of experiments 4(c) and 6 (c) described in reference (9).

IV EXPERIMENTAL RESULTS AND ANALYSIS

The diodes fabricated for this experiment are far from ideal.[10] There exists an interfacial layer between Au and GaAs. The interfacial

layer can be treated as a resistor in series with the Schottky diode.[10] Then the measured capacitance C' is[1]

$$C' = \frac{C}{1 + \omega^2 \gamma^2 C^2} \tag{7}$$

where is the equivalent series resistance. Since $\omega^2\gamma^2C^2 < 1$ in our experiments, Eqs. (2) and (6) still can be used to analyze the C-V data without introducing any appreciable error (i.e., $C \cong C'$).

The transient dark-capacitance measurements were performed between 285 and 316°K. A typical capacitance ΔC (i.e., $C(t=\infty)-C(t=0)$) versus time t curve is illustrated in Fig. 1. By using the Hewlett-Packard calculator 9100A and the least square curve fitting technique, the time constant (or the reciprocal of the thermal emission rate of electrons) was determined from Eqs. (2) and (6). The values of C(t=0) and $C(t=\infty)$ were also obtained.

A. Field Dependence of Thermal Emission Rate of Electrons

The thermal emission rate of electrons e_n as a function of the average electric field E is shown in Fig. 2 for several devices. At low electric field, the thermal emission rates remain constant for all devices. At higher electric field, the thermal emission rate of Device No. 17 (D-17) shows rapid increase while that of others remains unchanged. This can be explained qualitatively by the Poole-Frenkel effect (field-assisted thermal ionization)[11].

An electron is bound to the deep level impurity atom by some potential which may either be Coulomb's attractive or neutral potential. When an electric field is applied, the effect on the impurity potential is to lower the barrier that the trapped electron must overcome in order to escape from the deep impurity atom. This then increases the thermal emission rate since it requires less energy to release the electron. It is reasonable to assume that the potential barrier lowering by electric field is more effective on the neutral type potential. This is because of the nature of loose bounding between electron and neutral atoms. The less field dependence of the experimental results of Device D-7 seems to indicate an impurity potential that is more of the Coulomb attractive type. The stronger field dependent thermal emission rate of D-17 favors a neutral type potential. Indeed, the deep impurity in D- 7 is the donor

type oxygen.[8] The bounding force between oxygen atom and electron is the Coulomb attractive type. While the deep impurity in Device D-17 is an acceptor type chromium,[8] the bounding potential is a neutral type. Our experimental results agreed with these arguments.

The thermal emission rates of electron determined by by transient photo-capacitance method using interband light (i.e. $h\nu > E_g$) are also included in Fig. 2. The slight difference in magnitudes from the transient dark-capacitance measurement could be due to small temperature variation between the two separate measurements.

B. Thermal Activation Energy of Deep Level Impurities

The temperature dependence of the thermal emission rate of electrons is illustrated in Fig. 3. The data were taken at a bias voltage where the emission rate is not field dependent. From Eq. (5), we can determine the thermal activation energy of the deep impurities by calculating the slope of $\ell n(\frac{e_n}{T^2})$ versus the 1/T curve (see Fig. 3). The results for three devices are

D-7	(O)	$E_T = 0.80$ eV
D-11	(O)	$E_T = 0.78$ eV
D-17	(Cr)	$E_T = 0.72$ eV

Hence, in this measurement we have concluded that the thermal activation energies for oxygen and chromium in GaAs are 0.79ev and 0.72ev from the conduction band edge respectively. These values are in good agreement with those obtained by the optical and Hall effect measurements;[8] they are 0.80ev and 0.73ev for oxygen and chromium respectively.

C. Determination of Shallow Donor and Deep Level Impurity Concentration:

The $C(t=0)^{-2}$ vs. V and $c(t=\infty)^{-2}$ vs. V are shown in Fig. 4 for devices under study. Using Eqs. (3) and (4), the slope of C^{-2} vs. V plot gives the shallow and deep impurity concentrations. The results are as follows.

For D-7(O)	$N_D+N_T = 3.1\times10^{15}\,cm^{-3}$,	$N_D = 9.2\times10^{14}\,cm^{-3}$
		$N_T = 2.18\times10^{15}\,cm^{-3}$
D-11(O)	$N_D+N_T = 8.3\times10^{15}\,cm^{-3}$,	$N_D = 2.64\times10^{15}\,cm^{-3}$
		$N_T = 5.66\times10^{15}\,cm^{-3}$
D-17(Cr)	$N_D-N_T = 5.2\times10^{14}\,cm^{-3}$,	$N_D = 2.55\times10^{15}\,cm^{-3}$
		$N_T = 2.03\times10^{15}\,cm^{-3}$

Comparing these values with those obtained by our Hall effect measurement on the bulk n type GaAs samples, they are in good agreement (see Table 1).

The intercept of the C^{-2} vs. V plot on the voltage axis should be the diffurion potential of the diode. However, the existence of a high resistance interfacial layer makes the apparent potential V_{bi}' higher than the intrinsic value V_{bi}[12,13]. It has been shown that[1]

$$V_{bi} = V_{bi}' - 2\gamma^2\omega^2 \left[\frac{d(1/C)^2}{dV}\right]^{-1} \tag{8}$$

From Fig. 4 and Eq. (8), the series resistance γ can be estimated (V_{bi}=0.9eV). They are $\gamma \cong 4k\Omega$ for D-7, $\gamma \cong 3.2k\Omega$ for D-11 and $\gamma \cong 10k\Omega$ for D-17. The values are much larger than those of the ideal devices. Nevertheless they have had no appreciable effect on determining the properties of deep impurity centers (see Eq. (7)).

V. CONCLUSIONS

The capacitance-voltage characteristics of Au-GaAs (n-type) Schottky Barrier diodes have been studied, with the presence of deep level impurities and interfacial layer.

The thermal emission rates of electrons from oxygen and chromium impurities have been deduced from the transient capacitance measurements. From the field dependence of these thermal emission rates, the donor nature of oxygen impurity and the acceptor nature of chromium impurity are clearly demonstrated. From the temperature dependence of thermal emission rates, the thermal activation energies are found to be 0.79ev and 0.72ev from the conduction band edge for oxygen and chromium respectively. These values are in good agreement with those obtained by optical and Hall effect measurements.[7] For comparison, the results from the previous authors are included in Table 2 along with the present result. The existence of an interfacial layer does not introduce appreciable errors in determining the properties of deep impurity centers.

References

1. A. M. Goodman, J. Appl. Phys., 34, 329 (1963).
2. R. Williams, J. Appl. Phys., 37, 3411 (1966).
3. Y. Furukawa and Y. Ishibashi, Japan J. Appl. Phys., 5, 837 (1966); also, Japan J. Appl. Phys., 6, 13 (1967).
4. J. C. Carballes and J. Lebailly, Solid State Comm., 6, 167 (1968).
5. R. R. Senechal and J. Basinski, J. Appl. Phys., 39, 4581 (1968).
6. G. H. Glover, IEEE Trans. Electron. Dev., ED-19, 138 (1972).
7. Y. Zohta, Appl. Phys. Letter, 17, 284 (1970).
8. S. M. Sze and J. C. Irvin, Solid-State Electron., 11, 599 (1968)
9. C. T. Sah, L. Forbes, L. L. Rosier and A. F. TAsch, Jr., Solid State Electron., 13, 759 (1970)
10. C. I. Huang and S. S. Li, IEEE Proc. Letter, 61, 488 (1973).
11. A. F. Tasch, Jr. and C. T. Sah, Phys. Rev. B., 1, 800 (1970).
12. C. R. Crowell and G. Z. Roberts, J. Appl. Phys., 40, 3726 (1969).
13. A. M. Cowley, J. Appl. Phys., 37, 3024 (1966).

Table 2: Summary of the Results for Au-GaAs (n type) Schottky Barrier Diodes Deduced from Transient and Photo-capacitance Measurements at 300°K

Author	Deep Level Impurity	E_T (eV)	e_n (sec^{-1})	S (cm^2)	Methods
General Quoted Values[8]	O	0.80	-	-	Hall Effect
	Cr	0.73	-	-	Optical
This paper	0	0.79	0.072	-	Transient Dark-and Photo-Capacitance Measurements
	Cr	0.72	0.065		
Zohta[7]	0	0.57	-	-	C - V
Carballes and Lebailly[4]	0	0.66	-	2.1×10^{-15}	C - V
Furukawa and Ishibashi[3]	0	0.7	0.06	-	C - V
Williams[2]	0	0.74	0.15	5×10^{-15}	C - V
Glover[7]	0	0.83	-	-	C - V
Senechal and Basinski[5]	0	0.9	-	$1.2\ 4 \times 10^{-15}$	C - V

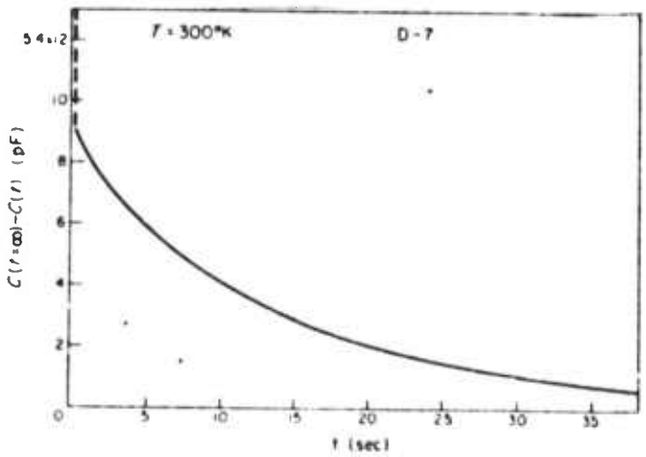


Fig. 1. The actual measured capacitance as a function of time.

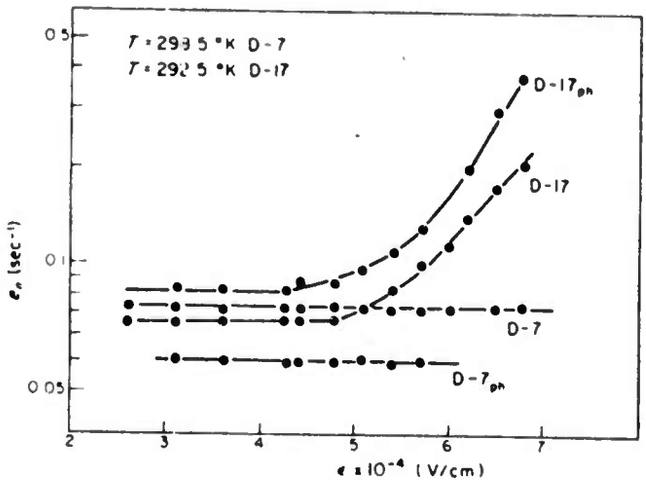


Fig. 2. The thermal emission rate of electron as a function of electric field. The subscription ph denotes that the transient photocapacitance method was being employed to obtain the data (λ = 0.6328 μm).

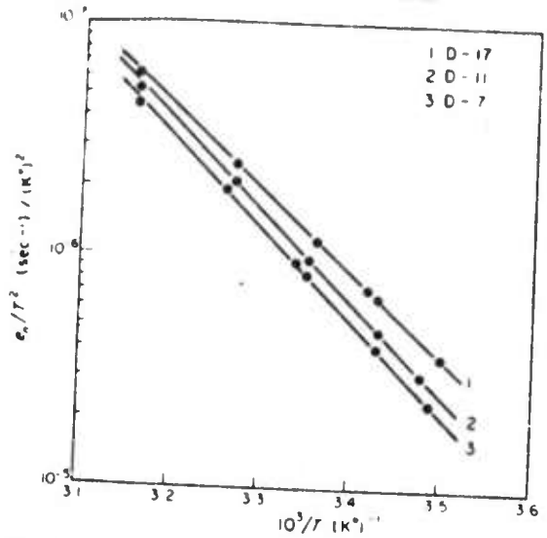


Fig. 3. The thermal emission rate of electron as a function of inverse temperature, $(10^3/T)$, between 285 and 316°K. The data were taken at $V_R = -6$ V.

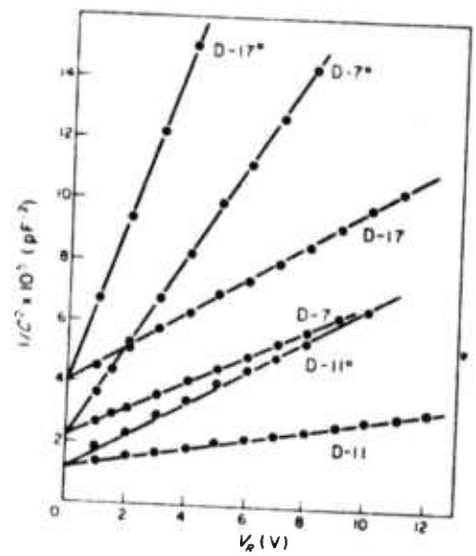


Fig. 4. A plot of C^{-2} as a function of reverse bias voltage for D-17, D-7 and D-11. The superscript * represents the values taken at $t = 0^+$. The unsuperscript represents the values taken at $t = \infty$.

VII. An Evaluation of Carrier Domain Devices for Functional Integrated Circuits (J. E. Smith and A. J. Brodersen)

The goal of this research has been to evaluate an innovative new concept in solid-state electronic devices, the carrier domain, which was introduced by Gilbert [1] [2] [3].

The principle of operation of a carrier domain device relies upon the creation of a limited region of injection within an elongated base-emitter structure of a bipolar junction transistor. The limited region of injection is achieved by the creation of a potential distribution within the base which exhibits a maximum or minimum value. By proper design of the base geometry and drive conditions the position of the potential maximum (minimum) can be controlled in a well defined manner by a differential input current to the base. Thus the limited region of injection in the base-emitter junction, a carrier domain, is analogous to a mobile elemental transistor. The carrier domain can be made to move along the length of the base-emitter junction above a thin resistive collector region which has two contacts and two external loads. Thus the injected current splits between the two loads. The differential collector output divided by the differential base input represents the device's transfer characteristic.

The first carrier domain device (CDD) introduced by Gilbert [1] was a four-quadrant, analog multiplier. The geometry of the device is shown in Fig. 1. For purposes of discussion, the half-disc of Fig. 2 will

be regarded as a two-dimensional resistive sheet which is approximated by either the upper or lower half of the base of the device shown in Figure 1. For the drive conditions shown in Fig. 2 it is shown in [4] by way of conformal mapping techniques that

$$\theta_m = \cos^{-1}(1-2\alpha)$$

where θ_m is the angle to the potential maximum along the perimeter of the half-disc, and α is the base-drive coefficient. Thus the device of Fig. 1 will exhibit potential maxima in its base at the angles θ_m. Injection into the collector will be greatest at $\theta = \theta_m$ and will fall to zero at angles remote to θ_m. The current splitting in the collector film will be proportional to X_m. Therefore since $X_m = \cos(\theta_m) = (1-2\alpha)$, Gilbert proposed that the transfer characteristic of either the upper or lower half of this device would be linear. The underlying assumption here is that the transfer characteristic is determined solely by the motion of the potential maximum within the base.

Gilbert's device accomplishes four-quadrant multiplication when the emitters and base regions are driven with differential current sources as shown in Fig. 1. To see how four-quadrant multiplier operation is achieved, first consider the conditions under which the differential collector current is zero. When $\alpha = 1/2$, $\theta_m = \pi/2$ and due to symmetry, the output is zero for all values of the emitter drive coefficient γ. Similarly, when the emitter drive is balanced ($\gamma = 1/2$) the output is zero for all values of α or θ_m. When both drive coefficients are perturbed away from the value 1/2, then the combination of the unbalanced positions and unbalanced magnitudes of the two carrier domains produces a differential collector output which is proportional to the product of the perturbations of α and γ from the value 1/2. This device is, in a

sense, a solid-state analog of a motor potentiometer multiplier. The linearity of the base-collector transfer characteristic is primarily determined by the motion of the carrier domains. The linearity of the emitter-collector transfer characteristic is determined by the common-base forward current gains of the inherent npn transistors. The CDD's promise of a base-collector transfer characteristic which is determined solely by the device's planar geometry offers the potential for the achievement of precision, linear and functional integrated circuits without sensitivity to variations in operating conditions or device processing while reducing the number of devices needed and the complexity of the circuit.

The extent to which the operation of a CDD can be determined by its planar geometry is a question of primary concern. A secondary but critically important question with regard to the overall usefulness of the carrier domain concept is whether or not a macroscopic two-dimensional representation is sufficient to describe the operation of the CDD. The theoretical and experimental studies which have been undertaken in order to answer these questions are treated in detail in [4] and are summarized below.

1) A base-emitter structure for carrier domain operation has been designed for which the two-dimensional potentials in the base region can be accurately described by a closed form mathematical analysis over a 95 percent dynamic range of the input base coefficient, α. The effects of the device's realistic contacts have been accurately accounted for. The design permits base drive current levels which are sufficient for carrier domain operation without introducing nonlinearities in the base region due to high electric field modulation of the mobility of the majority

carriers or reverse breakdown of the emitter-base junction. The emitter design avoids premature domain distortion at the ends of the emitter by permitting a 100 percent excursion of the emitter current density envelope.

2) The distribution and motion of the emitter current density envelope (the carrier domain) have been theoretically described in the closed form mathematical expressions. Attention has been focused upon the centroid of the domain and the two-dimensional motion of the centroid has been theoretically predicted by mathematical analysis.

3) The new base-emitter structure has been incorporated into the designs of two new CDD's with different collector structures. The two collector designs permit experimental measurement of the motion of the domain centroid along two different coordinate axes. Therefore these collector designs permit the two-dimensional motion of the domain centroid within the base-emitter structure to be completely determined experimentally.

4) The devices described above have been fabricated and their dc operation has been characterized. Study and comparison of the theoretically predicted and experimentally observed characteristics have revealed the limitations upon the carrier domain concept which must be accounted for in the design and application of CDD's.

The agreement obtained between the theoretically predicted and experimentally measured base-emitter and base-collector transfer characteristics of the devices studied indicates that a macroscopic, two-dimensional representation is entirely sufficient for an accurate description of the devices' terminal characteristics. On the basis of the results obtained for these specific

devices it is proposed that a macroscopic, two-dimensional representation is adequate for the analysis/design of any CDD which exhibits vertical dimensions which are uniform and are small compared to the planar dimensions of the base and collector regions of the device.

The selection of a collector design which accomplishes current division which is linearly dependent upon a single coordinate axis permits the adoption of the motion of the domain centroid as a conceptual link between the base-emitter and collector structures. Consequently the operation of the base-emitter and collector structures can be divorced and treated separately. This decomposition of the operation of the two major regions of the device greatly simplifies visualization of the operation of the CDD and permits independent analysis/design of the two regions.

Because the most judicious choice for a collector design is one in which current division is linearly dependent upon a single coordinate axis, a uniformly distributed resistive collector region is dictated. A uniformly distributed collector resistance can be effectively achieved by always incorporating a low resistivity buried layer under a higher resistivity epitaxial collector film. The resistivities of the buried and epitaxial layers and the thickness of the epitaxial layer must be carefully selected to provide shunting of the epitaxial collector film while avoiding base-substrate punch-through. Two planar geometries for the collector design which can perform the desired linear current splitting are: 1) a rectangular section with parallel stripe contacts; and 2) a semiannular section with radially parallel stripe contacts.

In view of the above observations it is concluded

that the operation of the collector region of a CDD and its load circuit do not pose any fundamental limitations upon the carrier domain concept.

A judicious selection of the design of the collector region permits the base-collector transfer characteristic to be completely determined by the motion of the domain centroid. Consequently the process of designing a CDD to achieve a specified base-collector transfer characteristic must focus upon the centroid motion. Independence of the base-collector transfer characteristic from operating conditions and processing variations can only be achieved with an emitter design which constrains the direction of the distribution and motion of the domain to lie along the same coordinate axis as the direction of current division in the collector. Futhermore, in order for the position of the domain centroid to be completely determined by the geometry of and current proportioning in the base region, the emitter current density distribution must exhibit symmetry about its maximum. The requirement of symmetry for the domain in turn requires that the potential in the base also exhibit a symmetrical distribution. Gilbert [1] [2] [3] has demonstrated that the above conditions are satisfied when the base structure exhibits parabolic attenuation. The domain motion which results in that case yields a linear base-collector transfer characteristic. However, no simple, uniformly distributed, continuous base geometry has been found for the implementation of the required parabolic attenuation [3].

In view of the above observations it is concluded that the carrier domain concept does indeed offer a mechanism for achieving high yield, high performance, linear and functional integrated circuits. However, the problem of finding geometric base designs which

simultaneously satisfy the requirements of domain symmetry and a specified motion of the potential maximum may prove to be difficult. It is recommended that future CDD related research efforts be directed toward solutions to the base design problem.

The CDD's base-emitter transfer characteristic offers a method of achieving non-linear function generation. The process of designing a CDD which is to exhibit a specified base-emitter transfer characteristic only needs to consider the magnitude of the potential maximum in the base region as a function of α. The emitter current density distribution and the collector region can be completely ignored. Furthermore, the magnitude of the potential maximum can be made relatively independent of operating and processing variables when the CDD and its drive circuitry are integrated on the same silicon chip. The base-emitter transfer characteristic will exhibit a translation dependence upon operating and processing variables by way of the dependence of the base-emitter junction potential's dependence upon these quantities. Consequently the base-emitter transfer characteristic will not be generally useful for non-linear function generation in integrated circuits.

BIBLIOGRAPHY

1. B. Gilbert, "New planar distributed devices based on a domain principle," 1971 IEEE International Solid-State Circuits Conference, Digest of Technical Papers, pp. 166-167.
2. ____, "Monolithic analog READ-ONLY memory for character generation," IEEE J. Solid-State Circuits, Vol. SC-6, pp. 45-55, Feb. 1971.
3. ____, Personal communication.
4. J. E. Smith, "Carrier Domain Electronics: A Spa-

tially Distributed Approach to Functional Integrated Circuits," Ph.D. Dissertation, University of Florida, 1974.

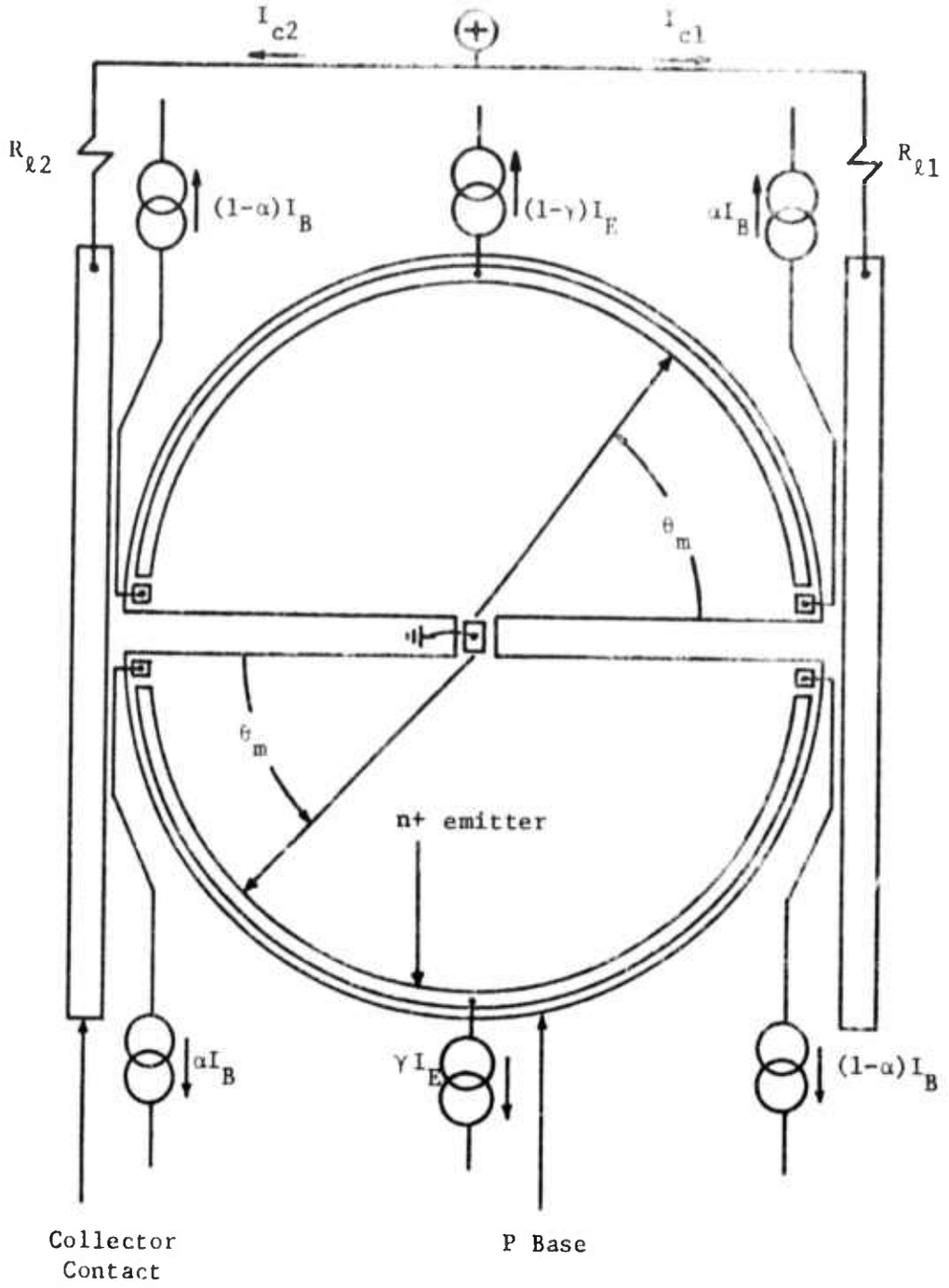


Figure 1. Planar geometry of an early carrier domain device introduced by Gilbert.

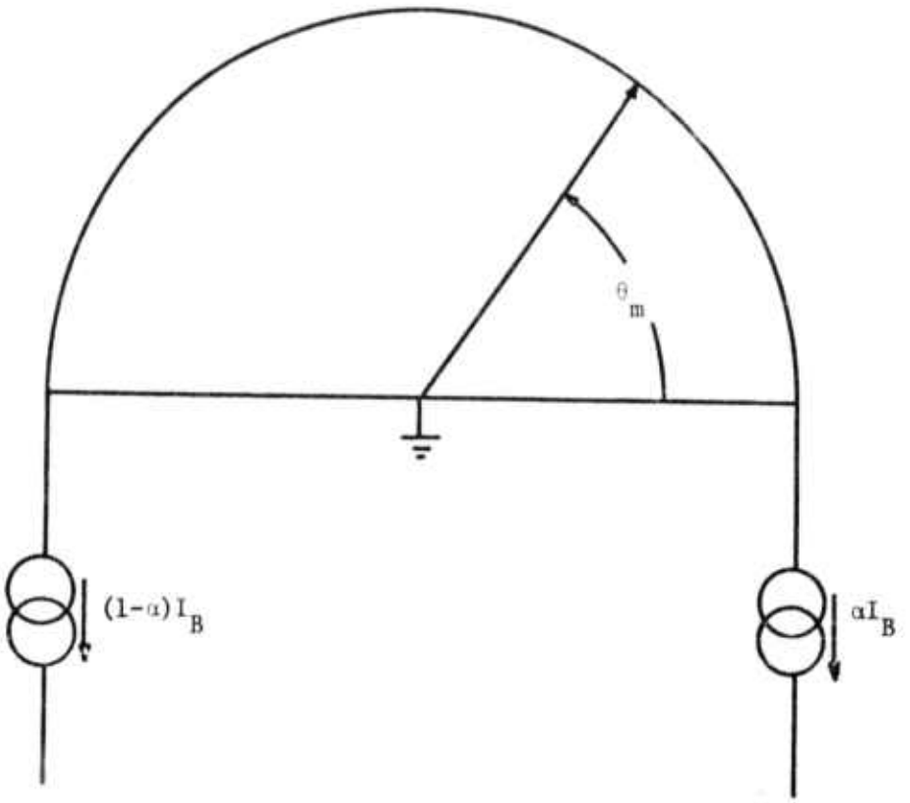


Figure 2. A two-dimensional resistive sheet with point contacts and ideal current sources.